

# **A high-performance reconstruction method combines total variation with a video denoiser for compressed ultrafast imaging**

Chengquan Pei<sup>1,5</sup>, David Day-Uei Li<sup>2</sup>, Qian Shen<sup>3</sup>, and Shian Zhang<sup>4</sup>, Dalong Qi<sup>4</sup>,  
Chengzhi Jin<sup>4,5</sup>, Le Dong<sup>1,6</sup>

1. Xidian University, Xi'an 710119, China
2. Faculty of Engineering, University of Strathclyde, Glasgow G4 0RE, Scotland, UK
3. Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
4. State Key Laboratory of Precision Spectroscopy, School of Physics and Materials Science, East China Normal University, 3663 North Zhongshan Road, Shanghai 200062, China
5. These authors contributed equally to this work
6. dongle@xidian.edu.cn

## **Abstract**

Compressed ultrafast photography (CUP) is a novel two-dimensional (2D) imaging technique to capture ultrafast dynamic scenes. Effective image reconstruction is essential in CUP systems. However, existing reconstruction algorithms mostly rely on image priors and complex parameter spaces. Therefore, in general, they are time-consuming and poor imaging quality, which limits their practical applications. In this paper, we proposed a novel reconstruction algorithm named plug-in-plug-Fast Deep Video Denoising Net-Total Variation (PnP-TV-FastDVDnet), which exploits images' spatial features and correlation features in the temporal dimension. Therefore, it offers higher-quality images than previously reported methods. Firstly, we built a forward mathematical model of the CUP, and the closed-form solution of the three sub-optimization problems was derived according to plug-in and plug-out frames. Secondly, we used an advanced video denoising algorithm based on a neural network named FastDVDnet to solve the denoising problem. Peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM) are improved on actual CUP data compared with traditional algorithms. On benchmark and real CUP datasets, the proposed method shows comparable visual results while reduces the running time by

96% over state-of-the-art algorithms.

**Key Words:** Compressed Ultrafast Photography, Compressed Sensing, Video Denoise

## 1. INTRODUCTION

Compressed ultrafast photography (CUP) [1] is a computational imaging technique that allows 2D imaging at picosecond or even femtosecond time scales. CUP combines compressed sensing (CS) [2] and streak camera technologies, capable of capturing irreducible transient events with only a single shot. Because of its high temporal resolution, it has significant potential to reveal underlying mechanisms in the physical, chemical and biological domains [2]. CUP is being applied widely in fluorescence lifetime detection [1, 5, 6], electronic imaging [7], and ultrafast laser dynamics [8, 9].

CUP can be divided into two steps [1]: the first step is image acquisition by streak cameras. This process can be viewed as multiple frames of video superimposed into a single observed image after random sampling and offsetting. The second step is data reconstruction, where the forward imaging model builds the inverse problem model. Then the observed images are converted into a video by an optimization algorithm. J. Hunt et al. derived theoretical guarantees for stable, accurate reconstruction at high compression ratios [10, 11]. As CUP reconstruction is an ill-posed problem, total variation (TV) [12-14], sparsity [15-18] and low rank [19, 20] are usually chosen as penalty terms in the optimization problem. The two-step iterative threshold method (TwIST) [21] is a classical CUP reconstruction algorithm that employs TV constraints in the reconstruction process [1, 8], and the algorithm has high efficiency but introduces artifacts. TwIST has other disadvantages, such as containing more artificially determined parameters, sensitive to initial values and challenging to converge. C. Yang utilized the Augmented Lagrangian (AL) algorithm to reduce the dependence on penalty term parameters and to obtain higher image reconstruction quality [22].

Recently Plug-and-Play (PnP) has been applied as a flexible and efficient algorithmic framework [23, 24] for solving inverse problems with convergence proofs [25, 26]. Its most significant advantage is that any state-of-the-art denoising algorithm

can be plugged in to improve the reconstruction quality. Y. Lai et al. decoupled the optimization variables [27] using the Alternating Direction Multiplier Method (ADMM) in the PnP framework. K. Dabov plugs in the denoising algorithm block-matching 3D filtering denoising (BM3D) [28] based on the PnP framework to solve the suboptimization problem. Although reconstructed images' quality can be improved, the reconstruction efficiency degrades since BM3D performs frame-wise and is time-consuming. With the development of Deep Learning (DL), some CNN-based image-denoising algorithms have better performance, such as the deep CNN denoiser prior for image restoration (IRCnn) [29], deep CNN for image denoising (DnCNN) [30], and fast and flexible solution for CNN-based image denoising (FFDNet) [31]. For scenarios (e.g., CUP) with insufficient data to support end-to-end training, PnP has the greater advantage of using a denoiser that is trained in advance [26]. This property allows CNN-based denoisers to be applied to the reconstruction process. Q. Shen applied the FFDNet to the algorithmic framework of PnP, and its reconstruction quality, speed, and flexibility exceed those of traditional reconstruction algorithms [32].

Image denoising algorithms are widely used denoisers in CUP reconstruction using the PnP framework. However, in the forward model, transient scenes that are continuous in time are discrete sampled into multiple frames, and there are correlations in the temporal dimension between frames. In previous work, the image denoising algorithm used in reconstruction based on the PnP can only exploit the spatial correlation of each frame. This paper applies the advanced CNN-based video denoising algorithm FastDVDNet [33] to the PnP to exploit the correlation between adjacent frames. The experimental and simulation results show that the proposed method improves reconstruction quality and efficiency.

## 2. Mathematical Model of CUP

The CUP system compresses multiple video frames into one frame. As shown in Fig. 1, the process of data acquisition contains three steps: Encoding, Shearing and Overlaying. A video dataset with  $N$  frames (the size =  $n_x \times n_y$ ) represents the ultrafast dynamic scene to be captured, randomly sampled by a mask (the size =  $n_x \times n_y$ ).

**In the encoding step**, the Hadamard (element-wise) product of the coding matrix

$C$  and dynamic scene  $X$  indicates that the encoding device, such as the DMD, randomly encodes each frame of the video data  $X_i$ . **In the shearing step**, under the action of the streak camera, each frame shears  $s_0$  pixel along the sweep direction from the previous frame, and this process can be represented by the shearing operator  $S_i$ . **In the overlaying step**, all images are accumulated into one observed image during the exposure time of the CCD embedded in the streak camera.

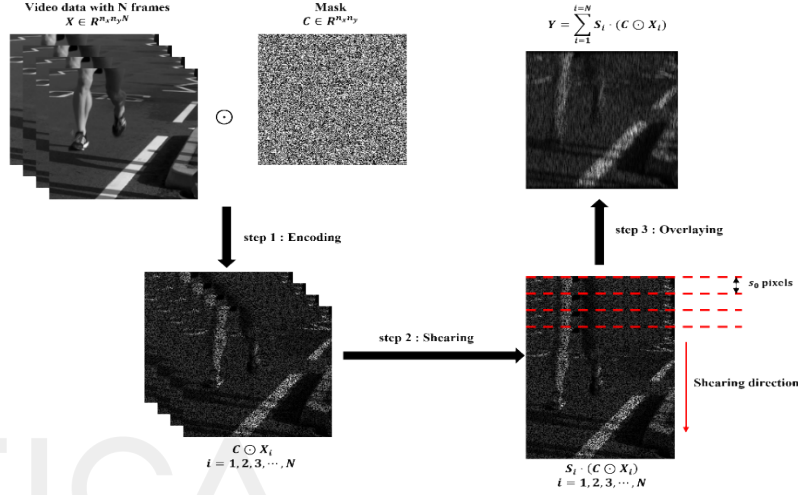


Fig.1 Data compression process of CUP

The CUP data compression process can be described as follows:

$$\mathbf{Y} = \sum_{i=1}^{i=N} \mathbf{S}_i \cdot (\mathbf{C} \odot \mathbf{X}_i) + \mathbf{Z}, \quad (1)$$

where the symbol  $\odot$  denotes the Hadamard product,  $\mathbf{Y}, \mathbf{Z} \in \mathbb{R}^{L \times n_y}$ ,  $L = n_x + (N-1)s_0$ , and  $Z$  indicates noise in the imaging process. To convert the Hadamard product into the standard matrix product, the 2D vectors  $\mathbf{X}_i$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$  in Eq. (1) need to be converted into one-dimensional vectors  $x_i$ ,  $y$  and  $z$ :

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^N \mathbf{S}'_i (\mathbf{C}_{diag} \mathbf{x}_i) + \mathbf{z} \\ &= \sum_{i=1}^N (\mathbf{S}'_i \mathbf{C}_{diag}) \mathbf{x}_i + \mathbf{z} \\ &= [\mathbf{S}'_1 \mathbf{C}_{diag}, \mathbf{S}'_2 \mathbf{C}_{diag}, \dots, \mathbf{S}'_N \mathbf{C}_{diag}] [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T]^T + \mathbf{z} \end{aligned} \quad (2)$$

Where,  $\mathbf{x}_i = \text{Vec}(\mathbf{X}_i) \in \mathbb{R}^{n_x n_y \times 1}$ ,  $\mathbf{y} = \text{Vec}(\mathbf{Y}) \in \mathbb{R}^{L n_y \times 1}$ , and  $\mathbf{z} = \text{Vec}(\mathbf{Z}) \in \mathbb{R}^{L n_y \times 1}$ . The symbol  $\text{Vec}(\bullet)$  indicates the transformation of a 2D vector into a 1D vector.  $\mathbf{C}_{diag}$  denotes the diagonal form of the encoding matrix  $\mathbf{C}$ ,  $\mathbf{C}_{diag} = \text{diag}(\text{Vec}(\mathbf{C})) \in \mathbb{R}^{n_x n_y \times n_x n_y}$ ,

and the symbol  $diag(\bullet)$  indicates the conversion of a 1D vector into a 2D diagonal matrix. Eq. (2) can be described as a classical linear inverse problem model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (3)$$

where  $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T]^T$ , the sensing matrix  $\mathbf{H} \in \mathbb{R}^{L n_y \times n_x n_y N}$  is a blocked diagonal matrix:

$$\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_N] = [\mathbf{S}'_1 \mathbf{C}_{diag}, \mathbf{S}'_2 \mathbf{C}_{diag}, \dots, \mathbf{S}'_N \mathbf{C}_{diag}], \quad (4)$$

where  $\mathbf{S}'_i = CircShift(\mathbf{I}_0, (i-1)s_0 n_y) \in \mathbb{R}^{L n_y \times n_x n_y}$ ,  $i = 1, 2, 3, \dots, N$ , the symbol  $CircShift(\mathbf{A}, l)$  indicates that the matrix  $\mathbf{A}$  is cyclically shifted by  $l$  units along the vertical direction.

### 3. Proposed reconstruction methods

#### 3.1 PnP-ADMM for CUP

For the linear inverse Eq. (3), the PnP-ADMM algorithm iteratively solves the following three problems to find the optimal solutions [34]:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \quad f(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{x} - \left( \mathbf{v}^k - \frac{1}{\rho} \mathbf{u}^k \right) \right\|_2^2, \quad (5)$$

$$\mathbf{v}^{(k+1)} = \mathcal{D}_\sigma \left( \mathbf{x}^{(k+1)} + \frac{1}{\rho} \mathbf{u}^{(k)} \right), \quad (6)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \rho(\mathbf{x}^{k+1} - \mathbf{v}^{k+1}), \quad (7)$$

For Eq. (6), as  $\mathbf{H}\mathbf{H}^\top = diag([\psi_1, \dots, \psi_{L n_y}]^\top)$  is a diagonal matrix, a closed-form solution can be obtained [32]:

$$\mathbf{x}^{k+1} = \left( \mathbf{v}^k - \frac{1}{\rho} \mathbf{u}^k \right) + \mathbf{H}^\top \left[ \mathbf{y} - \mathbf{H} \left( \mathbf{v}^k - \frac{1}{\rho} \mathbf{u}^k \right) \right] ./ (s + \rho \mathbf{I}), \quad (8)$$

where  $s = [\psi_1, \psi_2, \dots, \psi_{L n_y}]^\top$  and  $\mathcal{D}(\cdot)$  in Eq. (7) denotes the denoising algorithm. Based on the PnP framework, any advanced denoising algorithm can solve Eq. (6).

#### 3.2 PnP-GAP for CUP

The solution steps of the PnP-GAP algorithm are as follows:

$$\mathbf{x}^{k+1} = \boldsymbol{\theta}^k + \mathbf{H}^\top (\mathbf{H}\mathbf{H}^\top)^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}^k), \quad (9)$$

$$\boldsymbol{\theta}^{k+1} = \mathcal{D}_\sigma(\mathbf{x}^{k+1}). \quad (10)$$

Similarly,  $\mathcal{D}(\cdot)$  represents the denoising algorithm. Since  $\mathbf{H}\mathbf{H}^\top = diag([\psi_1, \dots, \psi_{L n_y}]^\top)$  is a diagonal matrix in CUP, Eq. (9) can be simplified as:

$$\mathbf{x}^{k+1} = \boldsymbol{\theta}^k + \mathbf{H}^\top (\mathbf{y} - \mathbf{H}\boldsymbol{\theta}^k) ./ \mathbf{s}, \quad (11)$$

where  $\mathbf{s} = [\psi_1, \psi_2, \dots, \psi_{L_{n_s}}]^\top$ .

### 3.3 Denoise operators

FastDVDNet is a two-stage structure, where 5 frames of images are divided into three groups as inputs to Block1 [33], and the outputs of the three Block1 are the inputs to Block2. The Architecture used in FastDVDnet can be found in ref [33]. Among them, three Block1 share parameters, and Block1 and Block2 have the same structure; it is a modified version of the U-Net network. It is worth noting that compared with the original U-Net, the network here has 2 down-sampling layers (whereas the original network has 4) with down-sampling achieved through Conv layers instead of pools. In addition, up-sampling is not attained through Bilinear interpolation or deconvolution but rather through Pixel-Shuffle. FastDVDNet has simpler structure than DVDNet. In image/video denoising, the L1 loss is commonly used as it keeps the overall information of the denoised image. However, we use the L<sub>2</sub> loss in FastDVDNet and the parameters of this denoise net are similar to those in ref [33].

## 4. Simulation Results

### a) Simulation on Benchmark Data

To verify the effectiveness of the proposed algorithm, we selected 8 frames from the benchmark dataset *runner, traffic, drop and crash* [35] and generated a compressed image according to the CUP data compression model. Then PnP-TV, PnP-BM3D, PnP-FFDNet, PnP-FastDvDNet and PnP-TV+FastDvDNet were used to recover video data of size 256×256×8 from the generated image of size 256×256, respectively. Besides, we also used the deep learning method (termed as DL) proposed in ref [36, 37], augmented Lagrangian and deep-learning hybrid algorithm (termed as AL+DL) proposed in ref [38] to compare the reconstruction performance. In addition, the spatial and intensity constraints (SIC) method (termed as TV+FastDvDNet-P) proposed in ref [39] was also used to provide a more stringent evaluation of the proposed method. In simulation, the streak camera measurement was generated according to the forward model. Similarly, the CCD measurement was generated by integrating the original dynamic scene along the time axis. The computing platform configuration parameters

we used are as follows: CPU is 12th Gen Intel(R) Core (TM) i7-12700H 2.70 GHz, GPU is NVIDIA GeForce RTX 3060 laptop GPU.

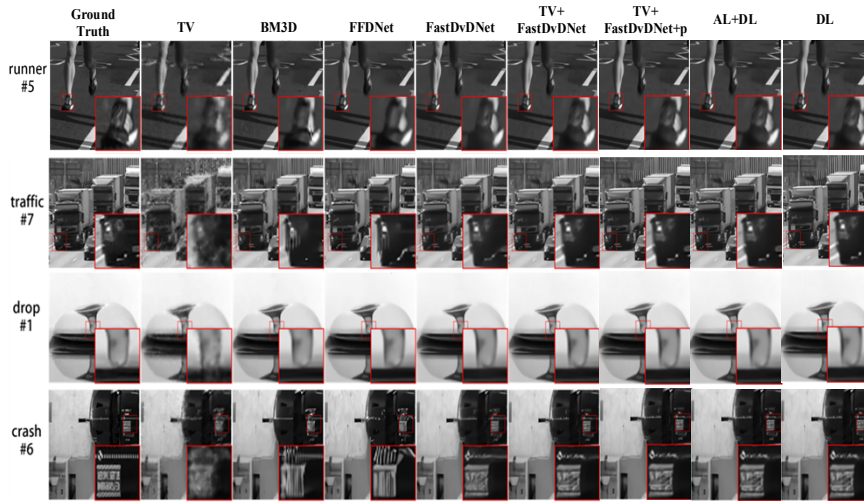


Fig. 2. Reconstruction results of algorithms using different denoisers

As shown in Fig. 2, one frame from the 8 frames reconstructed from each simulation dataset is selected for comparison. Artifacts exist in the reconstructed image of PnP-TV, and blurring occurs in the detail part, so the image's visual effect is relatively poor. Notably, the reconstruction performance of these algorithms is close when recovering *drop* data with a simple image structure. In contrast, the reconstruction results of PnP-BM3D and PnP-FFDNet show significant distortion when recovering *traffic* and *crash* data with complex image structures. Therefore, PnP-FastDvDNet and PnP-TV+FastDvDNet outperformed other algorithms in restoring image details.

Tables 1, 2 show the PSNR, SSIM of the reconstructed results, respectively. According to the average PSNR and SSIM, PnP-FastDvDNet is the best algorithm using a single denoiser, and the algorithm using the combined TV and FastDvDNet denoiser has 1.2 dB and 0.012 improvements in PSNR and SSIM respectively compared to it. Regarding algorithm execution efficiency, the algorithm using the BM3D denoiser takes the longest average time, while the algorithm using the TV and FFDNet denoiser takes less time. Still, the reconstruction is less effective than the FastDvDNet and TV+FastDvDNet denoiser algorithm. The execution time of PnP-TV+FastDvDNet is less than that of PnP-FastDvDNet because a part of its iterations uses a TV denoiser. However, PnP-TV+FastDvDNet works better due to the utilization

of both TV and video depth during reconstruction. The PSNR and SSIM were shown in Table 1 and Table 2. We also made an efficiency comparison for different methods as shown in Table 3.

Table 1 Average PSNR (dB)

Algorithm	Runner	Traffic	Drop	Crash	Average
AL+DL	32.12	24.56	38.51	25.50	30.1725
DL	31.24	23.55	38.01	24.91	29.7524
PnP-ADMM-TV	26.87	20.54	30.88	24.44	25.964
PnP- ADMM-BM3D	32.48	24.24	<b>40.48</b>	26.05	30.958
PnP- ADMM-FFDNet	31.88	23.22	38.59	24.58	30.122
PnP-ADMM-FastDvDNet	33.42	26.58	37.76	27.22	31.28
PnP-ADMM-TV+FastDvDNet	<b>34.81</b>	<b>27.51</b>	39.15	<b>27.25</b>	<b>32.446</b>
PnP-ADMM-TV+FastDvDNet-P	<b>35.01</b>	<b>28.12</b>	39.89	<b>28.35</b>	<b>32.842</b>
PnP-GAP-TV	28.3	25.4	35.4	26.5	28.57
PnP-GAP- BM3D	33.5	26.2	<b>39.1</b>	25.6	31.06
PnP-GAP- FFDNet	34.1	27.3	38.8	27.1	29.88
PnP-GAP-TV+FastDvDNet	<b>35.6</b>	<b>28.2</b>	38.08	<b>28</b>	<b>32.71</b>
PnP-GAP-TV+FastDvDNet-P	<b>35.89</b>	<b>28.98</b>	38.91	<b>28.56</b>	<b>33.01</b>

Table 2. Average SSIM

Algorithm	Runner	Traffic	Drop	Crash	Average
AL+DL	0.965	0.861	0.995	0.901	0.9305
DL	0.95	0.843	0.989	0.896	0.9291
PnP-ADMM- TV	0.905	0.729	0.953	0.863	0.8714
PnP-ADMM -BM3D	0.965	0.873	<b>0.994</b>	0.908	0.9402
PnP-ADMM -FFDNet	0.962	0.859	0.992	0.889	0.9326
PnP-ADMM-FastDvDNet	0.966	0.914	0.987	0.921	0.9474
PnP-ADMM-TV+FastDvDNet	<b>0.973</b>	<b>0.932</b>	0.989	<b>0.931</b>	<b>0.9582</b>
PnP-ADMM-V+FastDvDNet-P	<b>0.981</b>	<b>0.95</b>	0.99	<b>0.95</b>	<b>0.97</b>
PnP-GAP-TV	0.914	0.812	0.95	0.791	0.8694
PnP-GAP- BM3D	0.973	0.861	0.993	0.812	0.919
PnP-GAP- FFDNet	0.976	0.88	0.984	0.881	0.9372
PnP-GAP-TV+FastDvDNet	<b>0.98</b>	<b>0.912</b>	<b>0.99</b>	<b>0.911</b>	<b>0.9688</b>
PnP-GAP-TV+FastDvDNet-P	<b>0.985</b>	<b>0.92</b>	<b>0.99</b>	<b>0.921</b>	<b>0.97</b>

Table 3 Results of average execution time(second)

Algorithm	Runner	Traffic	Drop	Crash	Average
PnP-TV	21	22	18	20	20.2
PnP-BM3D	731	706	727	713	715.4
PnP-FFDNet*	26	27	25	26	26.2
PnP-FastDvDNet*	72	75	69	70	70.6
PnP-TV+FastDvDNet*	69	63	56	60	61.6
AL+DL	51	55	60	58	56
DL	0.1	0.1	0.1	0.1	0.1

Note: The symbol \* indicates the use of GPU

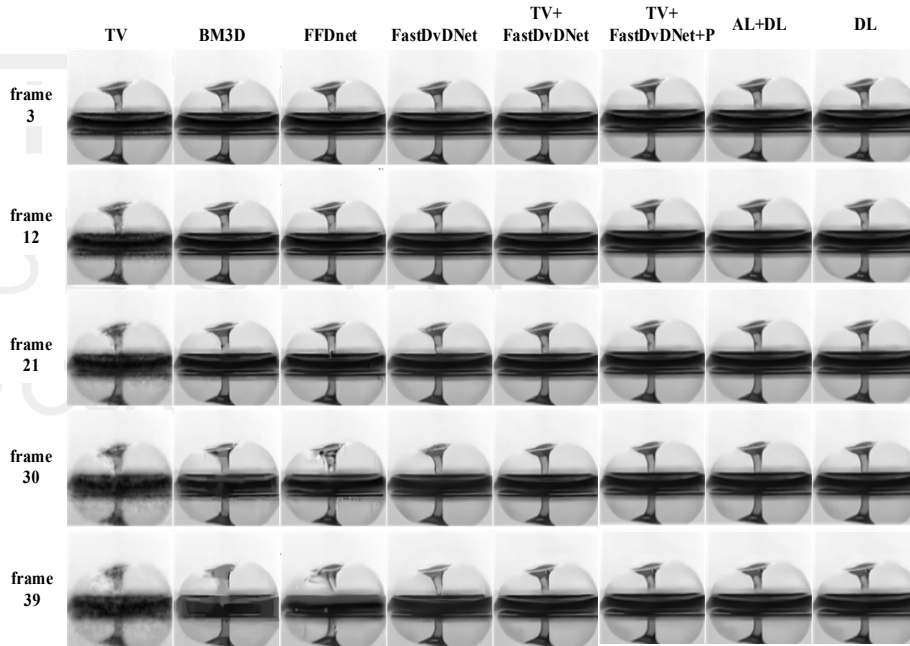
By applying the joint TV and FastDvDNet denoiser to the PnP, temporal correlations are exploited in the reconstruction process, and the reconstruction effect exceeds that of the algorithm using a single denoiser. At the same time, its execution time is less than that of PnP-FastDvDNet.

**b) Simulation on High Compression Ratio Data**

To test the reconstruction performance of the proposed algorithm on data with larger compression ratios (CR), 13 simulated datasets were generated by selecting 3, 6, 9, ..., and 39 frames from the *drop* dataset. The CUP data compression ratio is defined as follows:

$$R = \frac{size(\mathbf{x})}{size(\mathbf{y})} = \frac{n_x n_y N}{L n_y} = \frac{n_x N}{L} = \frac{n_x N}{n_x + s_0(N - 1)}. \tag{12}$$

The number of compressed frames positively correlates with the data CR.



**Fig. 3. Reconstruction results for different compression ratios**

Fig. 3 shows the reconstruction results when the number of compressed frames is 3, 12, 21, 30 and 39, respectively. Compared with PnP-TV+FastDvDNet, the algorithm using a single denoiser shows distortion and blurring in some areas of the reconstructed image when the CR increases. The results shows that PnP-TV+FastDvDNet delivers high reconstruction performance in terms of large CR.

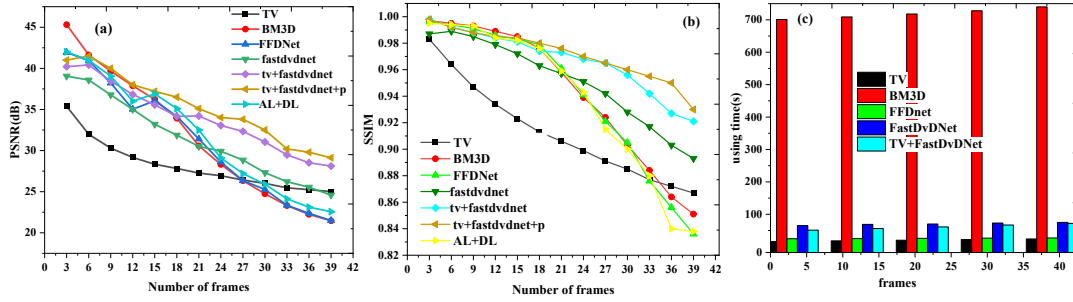


Fig.4 Graphs of (a) PSNR (dB), (b) SSIM and (c) using time with increasing compression ratio for reconstruction results

As shown in Fig.4, the PSNR, SSIM and using time of the reconstruction results using a single denoiser decrease rapidly when the CR increases. In contrast, the PnP-TV+FastDvDNet curve decreases more gently, and the reconstruction performance is better than other algorithms after the number of compressed frames exceeds 24 frames. Simulation results on data with high compression ratios show that our proposed method outperforms the algorithm using a single denoiser and also takes less time to reconstruct than PnP-FastDvDNet.

### c) Simulation on Different noise operators

With PnP we can apply advanced denoisers to the solutions of Eq. (3). Image-denoising algorithms are widely used in PnP-based optimization algorithms. However, in a CUP reconstruction problem, the image-denoising algorithms only recognize spatial features; they do not exploit the correlation between different video frames. For ultrafast dynamic scenes, considering the timing correlation between sequential frames can guarantee higher-quality reconstructions.

In contrast, our FastDvDNet denoising considers image spatial features and the temporal correlation between video frames. We integrate TV and FastDvDNet denoising with PnP, using both TV and video deep prior, to improve image reconstruction quality.

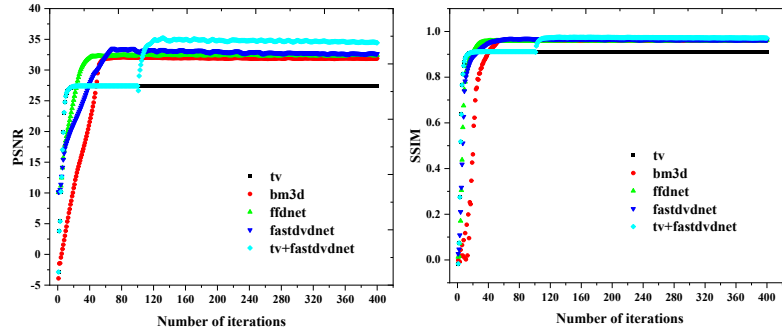


Fig.5 (a) The PSNR (dB) variation with the number of iterations. (b) The SSIM variation with the number of iterations.

As shown in Fig. 5, we applied different denoisers to the reconstruction process. The optimization algorithm has iterated a total of 450 times. For PnP-TV+FastDvDNet, the TV denoiser was used for 225 iterations, followed by the FastDvDNet denoiser for 225. All other methods use a single denoiser iterating 150 times. Compared with TV-only or FastDvDNet-only denoisers, the integrated TV and FastDvDNet denoiser shows better PSNR and SSIM performances, even surpassing the image denoiser BM3D and FFDNet.

#### d) Balance between the TV and FastDVDnet

In order to reveal the principle of denoising, the total denoising expression can be described as  $\lambda_1 \text{FastDvDnet} + \lambda_2 \text{TV}$ ,  $\lambda_2 = k\lambda_1$ . We set  $\lambda_1 = 0.0001$ , and calculated the PSNR when  $k$  is 1, 2, 3, ..., 10, respectively, and the result was shown in Fig.6.

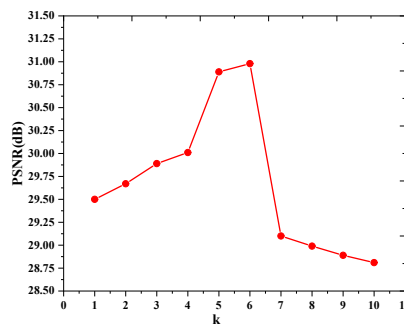


Fig.6 The PSNR curves as functions of the parameters  $\lambda_1$  and  $\lambda_2$

From Fig.6, we can see that the performance of the proposed method is insensitive to the variation in the values of  $\lambda_2$  in the range of 0.0001~0.001.

## 5. Real Data

As indicated in Fig. 7, an ultrafast dynamic scene is collected by an objective lens (OL) and imaged onto a DMD (Texas Instruments, DLP Light Crafter 3000), loading a

static random pattern.

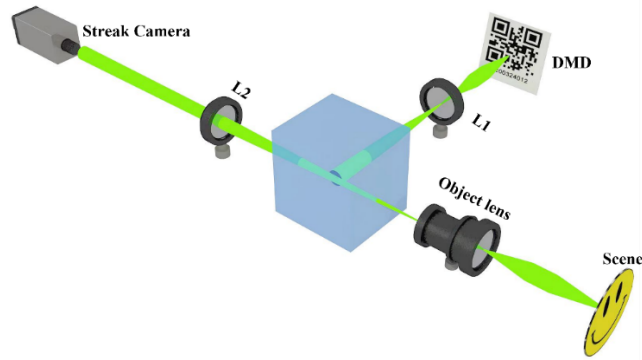


Fig. 7 Experimental system setup

After that, the encoded ultrafast scene of each channel is transferred through a  $4f$  imaging system composed of  $L_1$  and  $L_2$  and reflected by a reflecting prism. Finally, all of them are captured by a streak camera (Hamamatsu, C7700) with the slit fully opened. This scientific instrument can capture ultrafast scenes based on the photo-electric effect and temporal shearing technique [33,34]. In the first scene, a single laser pulse from a mode-locked Ti: Sapphire laser amplifier (Spectra-Physics, 50 fs) was stretched up to 200 ps by a pulse stretcher. The stretched pulse was spatially expanded to illuminate hollow letters 'E' and 'N' fabricated in a black nylon plate. Photons inside the shape could pass through the plate, while those outside were blocked. The resulting E-shaped laser pulse was projected onto a thin white paper sheet for scattering and further observed by the CUP system. Fig.8 shows the compressed data  $E$  and  $N$  captured by the CUP system. Data  $E$  and  $N$  sequence depths are 21 and 30, respectively.

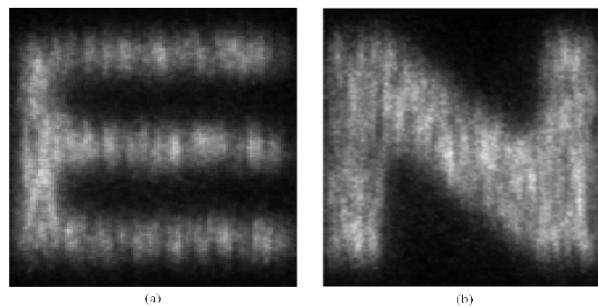


Fig. 8. Simply CUP real data E and N.

We performed the proposed method, and Figs 9 and 10 show the reconstructed outcomes.

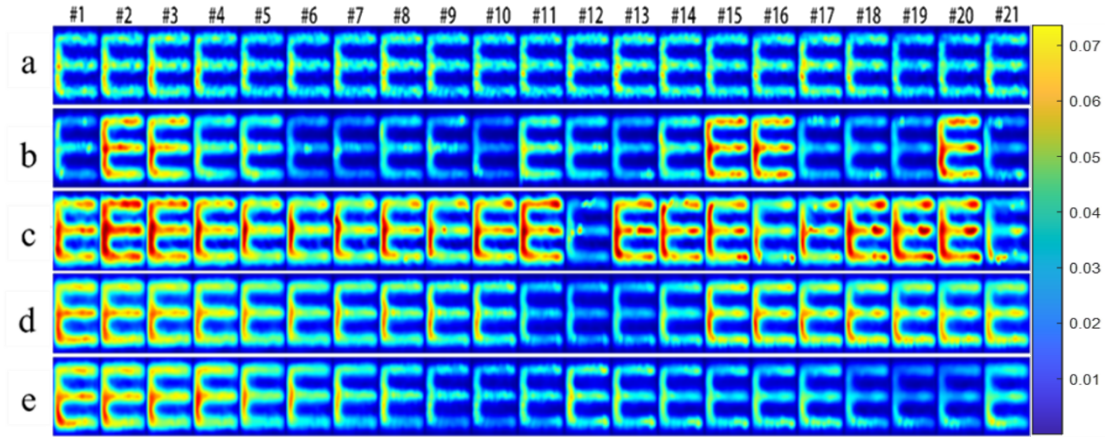


Fig. 9 Reconstruction results of real data  $E$  and the comparison among (a)TV, (b) BM3D, (c)FFDnet, (d)FastDVDnet, (e) TV+ FastDVDnet

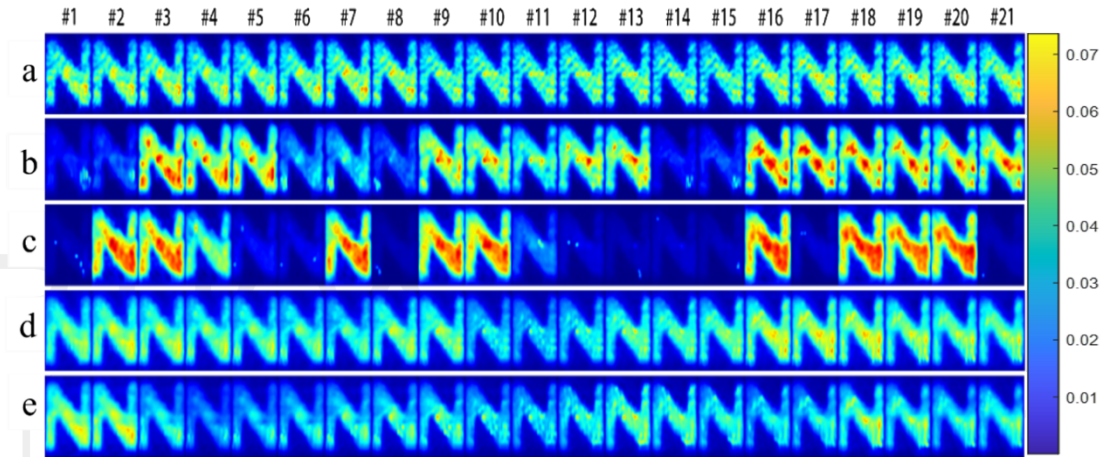


Fig. 10 Reconstruction results of real data  $N$  and the comparison among (a)TV, (b) BM3D, (c)FFDnet, (d)FastDVDnet, (e) TV+ FastDVDnet

Fig. 9 and 10 show the reconstruction results for real data  $E$  and  $N$ , respectively. Intensity variations are difficult to observe between different reconstruction frames for PnP-TV, whereas intensity variations can be observed but are not continuous for PnP-BM3D and PnP-FFDNet. FastDvDNet's reconstruction result shows a more coherent intensity change between adjacent frames, guaranteeing a better visual effect.

Light springs (LSs), as a type of novel-shaped ultrafast laser beam carrying orbital angular momentum (OAM), have a helical structure in both phase and intensity profile [40]. In order to extend the experiments to more complex real-world data. We conduct an experiment to capture the LSs, and the experiment details can be found in ref [40]. In above simulations and experiments, we have been demonstrated that the TV+FastDvDNet has advantage over state-of-the-art algorithms. Therefore, we

reconstruct the compressed image only using FFDnet, FastDVDnet, TV+ FastDVDnet to manifest the efficiency of proposed method. The reconstructed result was shown in Fig.11.

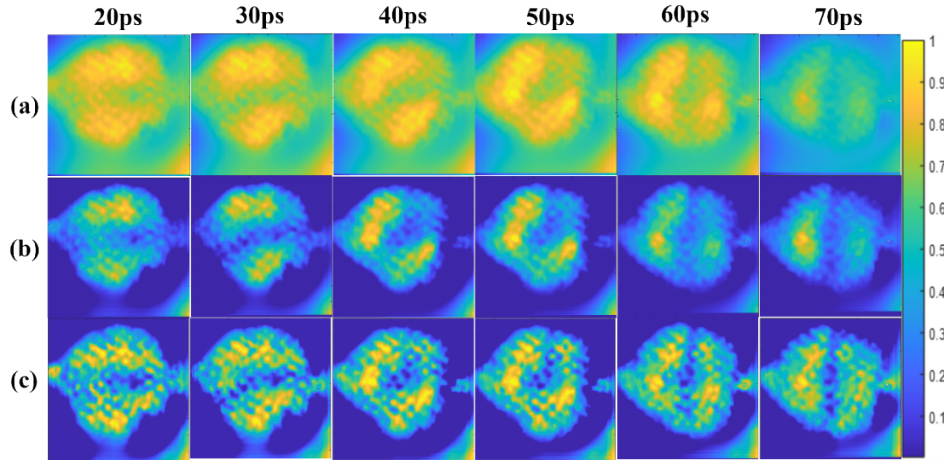


Fig.11 Reconstruction results of LSs and the comparison among (a)FFDnet, (b) FastDVDnet, (c)TV+ FastDVDnet

As shown in Fig.11, we can see that TV+ FastDvDNet’s reconstruction result shows a more coherent intensity change between adjacent frames, guaranteeing a better visual effect. To complement the 1D evolution information of the reconstructed data, the initial intensity integral information is supplemented. As demonstrated in Fig.12, a comparison reveals that the temporal evolution characteristics of the reconstructed data closely resemble those of the original data.

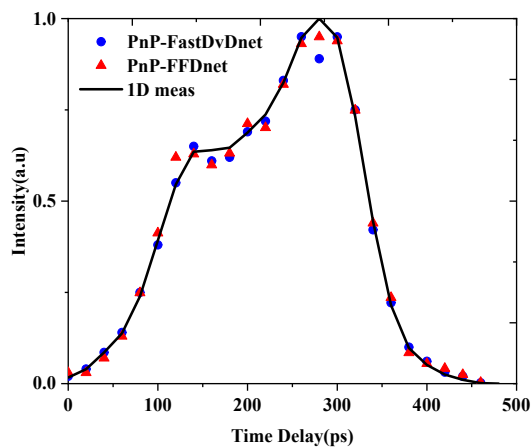


Fig.12 Curve of the temporal evolution of the laser pulse intensity

Information of the temporal evolution of the laser pulse intensity is also extracted for further comparison and shown in Fig.12. The black line for reference indicates the time evolution of the pulse intensity measured by the one-dimensional (1D) streak

camera with its entrance slit being narrowed to about 50  $\mu\text{m}$  for spatial sampling. Fig. 12 shows that all the data extracted from reconstructed images are highly agree well with the reference, so it manifests that our methods are capable for image reconstruction in CUP.

## 6. Conclusions

The proposed PnP-TV+FastDvDNet exploits both images' TV and the deep video prior, offering a higher quality reconstruction. Reconstruction results on simulated and real data show that using the integrated TV and FastDvDNet denoiser outperforms single-denoiser algorithms. Since the sequence depth of current CUP systems has exceeded 1000 frames, therefore, the proposed method provides much better reconstruction with a high sequence depth.

### Funding

National Natural Science Foundation of China (62101302);

### Disclosures.

The authors declare no conflicts of interest.

### References

- [1] L. Gao, J. Liang, L. Lin, and L. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," *Nature*, vol. 516 (74-77), 2014.
- [2] Y. C. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications," Cambridge University Press, 2012.
- [3] J. Liang and L. V. Wang, "Single-shot ultrafast optical imaging," *Optica*, 5(9), 2018.
- [4] D. Qi et al., "Single-shot compressed ultrafast photography: a review," *Adv. Photonics*, vol. 2, p. 014003 - 014003, 2020.
- [5] P. Wang, J. Liang, and L. V. Wang, "Single-shot ultrafast imaging attaining 70 trillion frames per second," *Nat. Commun.*, vol. 11, 2020.
- [6] J. V. Thompson, J. D. Mason, H. T. Beier, and J. N. Bixler, "High speed fluorescence imaging with compressed ultrafast photography," presented at, 2017.
- [7] X. Liu, S. Zhang, A. Yurtsever, and J. Liang, "Single-shot real-time sub-nanosecond electron imaging aided by compressed sensing: Analytical modeling and simulation.," *Micron*, vol. 117, p., 2019.
- [8] J. Liang, L. Zhu, and L. V. Wang, "Single-shot real-time femtosecond imaging of temporal focusing," *Light, Science & Applications*, vol. 7, 2018.
- [9] T. Kim, J. Liang, L. Zhu, and L. V. Wang, "Picosecond-resolution phase-sensitive imaging of transparent objects in a single shot," *Sci. Adv.*, vol. 6, 2020.

- [10] E. J. C. E. S, "The restricted isometry property and its implications for compressed sensing," *C. R. Math.*, vol. 346, pp. 589-592, 2008.
- [11] J. Hunt et al., "Metamaterial Apertures for Computational Imaging," *Science*, vol. 339, pp. 310 - 313, 2013.
- [12] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An Iterative Regularization Method for Total Variation-Based Image Restoration," *Multiscale Model. Simul.*, vol. 4, pp. 460-489, 2005.
- [13] E. Herrholz and G. Teschke, "Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems," *Inverse Probl.*, vol. 26, p. 125012, (2010).
- [14] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," 2016 IEEE International Conference on Image Processing (ICIP), pp. 2539-2543, (2015).
- [15] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust Single Image Super-Resolution via Deep Networks With Sparse Prior," *IEEE Trans. Image Process.*, 25(7) : 3194-3207,(2016).
- [16] P. Llull et al., "Coded aperture compressive temporal imaging," *Opt. Express*, 21(9): 10526-10545(2013).
- [17] M. Iordache, J. E. M. Bioucas-Dias, and A. J. Plaza, "Total Variation Spatial Regularization for Sparse Hyperspectral Unmixing," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, pp. 4484-4502, 2012.
- [18] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," *CVPR 2011*, pp. 329-336, 2011.
- [19] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M. Yang, "Image Deblurring via Enhanced Low-Rank Prior," *IEEE Trans. Image Process.*, vol. 25, pp. 3426-3437, 2016.
- [20] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted Nuclear Norm Minimization with Application to Image Denoising," 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2862-2869, 2014.
- [21] J. M. Bioucas-Dias and M. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992-3004, 2008.
- [22] C. Yang et al., "Improving the image reconstruction quality of compressed ultrafast photography via an augmented Lagrangian algorithm," *J. Opt.*, vol. 21, 2019.
- [23] S. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play priors for model based reconstruction," 2013 IEEE Global Conference on Signal and Information Processing, pp. 945-948, 2013.
- [24] S. Sreehari et al., "Plug-and-Play Priors for Bright Field Electron Tomography and Sparse Interpolation," *IEEE Trans. Comput. Imaging*, vol. 2, pp. 408-423, 2015.
- [25] S. H. Chan, X. Wang, and O. A. Elgendy, ""Plug-and-Play ADMM for Image Restoration: Fixed-Point Convergence and Applications," *IEEE Trans. Comput. Imaging*, vol. 3, pp. 84-98, 2016.
- [26] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-Play Methods Provably Converge with Properly Trained Denoisers," presented at, 2019.

- [27] Y. Lai et al., "Single - Shot Ultraviolet Compressed Ultrafast Photography," *Laser & Photonics Reviews*, vol. 14, 2020.
- [28] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080-2095, 2007.
- [29] Z. Kai, W. Zuo, S. Gu, and Z. Lei, "Learning Deep CNN Denoiser Prior for Image Restoration," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [30] K. Ma et al., "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004--1016, 2016.
- [31] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising," *IEEE Trans. Image Process.*, p. 1-1, 2017.
- [32] Q. Shen, J. Tian, and C. Pei, "A Novel Reconstruction Algorithm with High Performance for Compressed Ultrafast Imaging," *Sensors (Basel, Switzerland)*, vol. 22, 2022.
- [33] M. Tassano, J. Delon, and T. Veit, "FastDVDnet: Towards Real-Time Deep Video Denoising Without Flow Estimation," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1351-1360, 2019.
- [34] S. H. Chan, X. Wang, and O. A. Elgandy, "Plug-and-Play ADMM for Image Restoration: Fixed-Point Convergence and Applications," *IEEE Trans. Comput. Imaging*, 2017.
- [35] X. Yuan, Y. Liu, J. Suo, and Q. Dai, "Plug-and-Play Algorithms for Large-Scale Snapshot Compressive Imaging," presented at, 2020.
- [36] Parker, Matthew, "CUP-Net: Compressed Ultrafast Photography Using Convolutional Neural Networks" (2020). ENGS 88 Honors Thesis (AB Students). 15.
- [37] Yayao Ma, Xiaohua Feng, and Liang Gao, "Deep-learning-based image reconstruction for compressed ultrafast photography," *Opt. Lett.* 45, 4400-4403 (2020).
- [38] Chengshuai Yang, Yunhua Yao, Chengzhi Jin, Dalong Qi, Fengyan Cao, Yilin He, Jiali Yao, Pengpeng Ding, Liang Gao, Tianqing Jia, Jinyang Liang, Zhenrong Sun, and Shian Zhang, "High-fidelity image reconstruction for compressed ultrafast photography via an augmented-Lagrangian and deep-learning hybrid algorithm," *Photon. Res.* 9, B30-B37 (2021).
- [39] Liren Zhu, Yujia Chen, Jinyang Liang, Qiaofeng Xu, Liang Gao, Cheng Ma, and Lihong V. Wang, "Space- and intensity-constrained reconstruction for compressed ultrafast photography," *Optica* 3, 694-697 (2016).
- [40] Jin, C., Qi, D., Yao, Y. et al. Single-shot real-time imaging of ultrafast light springs. *Sci. China Phys. Mech. Astron.* 64, 124212 (2021).