PHOTONICS Research

High-fidelity image reconstruction for compressed ultrafast photography via an augmented-Lagrangian and deep-learning hybrid algorithm

Chengshuai Yang,[1] Yunhua Yao,[1,6] (ID) Chengzhi Jin,[1] Dalong Qi,[1] Fengyan Cao,[1] Yilin He,[1] Jiali Yao,[1] Pengpeng Ding,[1] Liang Gao,[2] Tianqing Jia,[1] Jinyang Liang,[3] (ID) Zhenrong Sun,[1] and Shian Zhang[1,4,5,7]

[1]State Key Laboratory of Precision Spectroscopy, School of Physics and Electronic Science, East China Normal University, Shanghai 200062, China
[2]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[3]Institut National de la Recherche Scientifique, Centre Énergie Matériaux Télécommunications, Laboratory of Applied Computational Imaging, Varennes, Québec J3X1S2, Canada
[4]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
[5]Collaborative Innovation Center of Light Manipulations and Applications, Shandong Normal University, Jinan 250358, China
[6]e-mail: yhyao@lps.ecnu.edu.cn
[7]e-mail: sazhang@phy.ecnu.edu.cn

Compressed ultrafast photography (CUP) is the fastest single-shot passive ultrafast optical imaging technique, which has shown to be a powerful tool in recording self-luminous or non-repeatable ultrafast phenomena. However, the low fidelity of image reconstruction based on the conventional augmented-Lagrangian (AL) and two-step iterative shrinkage/thresholding (TwIST) algorithms greatly prevents practical applications of CUP, especially for those ultrafast phenomena that need high spatial resolution. Here, we develop a novel AL and deep-learning (DL) hybrid (i.e., AL + DL) algorithm to realize high-fidelity image reconstruction for CUP. The AL + DL algorithm not only optimizes the sparse domain and relevant iteration parameters via learning the dataset but also simplifies the mathematical architecture, so it greatly improves the image reconstruction accuracy. Our theoretical simulation and experimental results validate the superior performance of the AL + DL algorithm in image fidelity over conventional AL and TwIST algorithms, where the peak signal-to-noise ratio and structural similarity index can be increased at least by 4 dB (9 dB) and 0.1 (0.05) for a complex (simple) dynamic scene, respectively. This study can promote the applications of CUP in related fields, and it will also enable a new strategy for recovering high-dimensional signals from low-dimensional detection.   © 2021 Chinese Laser Press

https://doi.org/10.1364/PRJ.410018

1. INTRODUCTION

Ultrafast imaging has played an indispensable role in photochemistry [1,2], biomedicine [3–5], microfluidics [6], shock waves [7], and plasma physics [8]. Recently, various ultrafast imaging techniques have been developed, including compressed ultrafast photography (CUP) [9–11]. Unlike some active ultrafast imaging techniques that need specific illumination light [12–14] or a pump–probe technique that requires multiple measurements [15–17], CUP is a single-shot and passive ultrafast imaging technique. Its temporal resolution and number of frames can reach tens of femtoseconds and several hundred, respectively. Therefore, CUP has great advantages for measuring some self-luminous or non-repeatable ultrafast phenomena, which is attributed mainly to the novel model of

CUP, which combines compressed sensing (CS) theory and time–space conversion technology. So far, CUP has been successfully applied to measure light reflection and refraction [9], femtosecond temporal focusing [10], photonic Mach cones [18], dissipative solitons [19], phase-sensitive transparent objects [20], three-dimensional (3D) objects [21], ultrashort laser spatiotemporal evolution [22], and photoluminescence processes [9]. However, due to the high data compression ratio, the fidelity of reconstructed images for CUP is relatively low by the conventional two-step iterative shrinkage/thresholding (TwIST) algorithm, which limits its practicality. To improve image fidelity, a variety of methods have been proposed, such as a space- and intensity-constrained image reconstruction algorithm [23], augmented-Lagrangian (AL)-based image

reconstruction algorithm [24], plug-and-play alternating direction method of multipliers algorithm [25], optimizing the codes for CUP [26], lossless CUP [18], and multi-encoding CUP [27]. These proposed schemes can improve image fidelity to a certain extent, but there are still great challenges in measuring the complex dynamic scenes.

In image reconstruction of CUP, all selections of the sparse domain, determination of relevant iteration parameters, and denoising after iteration calculation greatly limit image fidelity. To completely solve these problems, we developed a novel image reconstruction method based on an AL and deep-learning (DL) hybrid (i.e., AL + DL) algorithm. This idea is borrowed mainly from some early algorithms, such as the AL algorithm [24,28,29], learning iteration parameters [30–33], learning sparse domain [34–37], and U-net architecture [38], but there are still many differences compared to each of the early algorithms. First, the AL + DL algorithm utilizes multiple learning transformations to seek the best sparse domain. Typically, the sparse domain in conventional TwIST and AL algorithms is determined before image reconstruction [24,39], so it is usually not optimal for one dynamic scene. In contrast, the sparse domain in the AL + DL algorithm can be optimized in multiple transformations, which is more pertinent. Second, the AL + DL algorithm takes full advantage of gradient descent (GD), DL, and AL algorithms, which simplifies the mathematical architecture to deal with the 3D tensor problem, and these advantages can reduce the cost of each iteration and decrease the number of iterations. Third, the AL + DL algorithm optimizes the relevant iteration parameters by learning the dataset, which is different from previous AL and TwIST algorithms, where these parameters are artificially predetermined. Finally, the AL + DL algorithm uses a U-net architecture containing attention layers to help denoise and retain the spatial details of the images after iteration calculation. Importantly, our theoretical simulation and experimental results show the AL + DL algorithm can obtain much higher image fidelity than conventional AL and TwIST algorithms for CUP, which strongly supports our theory.

2. PRINCIPLE

In CUP, a 3D dynamic scene $I(x, y, t)$ is encoded by operator C, sheared by operator S, and integrated by operator T, and finally a two-dimensional (2D) image $E(x', y')$ is obtained. For convenience, hereafter, $I(x, y, t)$ is abbreviated to I, and $E(x', y')$ is abbreviated to E. Mathematically, this process can be described as

$$E = TSCI. \tag{1}$$

For simplicity, we define $O = TSC$. Thus, Eq. (1) can be further written as

$$E = OI. \tag{2}$$

To recover 3D I from 2D E, we need to solve the inverse problem of Eq. (2). The number of elements in I is much larger than that in E, so the inverse problem of Eq. (2) is undetermined. The CUP strategy is to introduce a CS theory [9]. The CS theory makes full use of the sparsity of I in a certain domain to recover the original information. This sparsity in one domain means that only a few elements are nonzero, while most of the

elements are zero. Consider a case in which I has n elements and E has m elements in the original domain, and I has s nonzero elements in a sparse domain, i.e., the sparsity where $n \gg s$ and $n > m > s$. Due to the fact that m is generally larger than s, this makes it possible to solve the inverse problem of Eq. (2). In a practical solution, the CS algorithm minimizes I in a sparse domain on condition of Eq. (1), which is shown as

$$\begin{cases} \min_I \Phi(I) \\ \text{s.t. } E - OI = 0 \end{cases}, \tag{3}$$

where $\Phi(I)$ is the expression of I in the sparse domain. According to CS theory [40,41], the original dynamic scene I can be completely recovered when

$$m > f s \mu^2, \tag{4}$$

where f is a constant correlated with the number of elements n, μ is the mutual coherence between the sparse basis of the sparse domain Φ and operator O, determined by operators C, S, and T. From Eq. (4), one can see that both increasing m and reducing s and μ are feasible schemes to improve the quality of image reconstruction. However, increasing m, i.e., increasing the sampling rate, will reduce the spatial resolution or requires many streak cameras, which is impractical in the actual CUP system. Thus, reducing s and μ is the best choice. Optimizing O can reduce only μ, while optimizing Φ can reduce both s and μ; therefore, here we employ the method of optimizing Φ. To optimize Φ, we impose a low-rank property on the entire dynamic scene I (tensor) with many different transformations [42], which is different from traditional methods with only one transformation. Thus, problem (3) can be further written as

$$\begin{cases} \min_{I(x, y, t)} \sum_p^q \psi_p I \\ \text{s.t. } E - OI = 0 \end{cases}, \tag{5}$$

where ψ_p represents one transformation in the sparse domain, and q denotes the total number of transformations. In transforming problem (5) from a constraint into an unconstraint, there exist two frameworks: the penalty function method and the AL method. The performance of the AL method is better than that of the penalty function method, which has been proved in previous works [24,29], and therefore, here the AL method is adopted. Thus, problem (5) can be transformed into

$$\min_I \left\{ \sum_p^q \psi_p I - \gamma(E - OI) + \frac{\zeta}{2} \|E - OI\|_F^2 \right\}, \tag{6}$$

where γ and ζ are the Lagrangian multiplier and penalty parameter, respectively, which are associated with I. For convenience, problem (6) is further written as

$$\min_I \left\{ \sum_p^q \psi_p I + \frac{\zeta}{2} \left\| E - \frac{\gamma}{\zeta} - OI \right\|_F^2 \right\}. \tag{7}$$

To solve problem (7), an auxiliary variable J is introduced into problem (7), and is written as

$$\begin{cases} \min_{I} \left\{ \sum_{p}^{q} \psi_p J + \frac{\zeta}{2} \left\| E - \frac{\gamma}{\zeta} - OI \right\|_F^2 \right\}. \\ \text{s.t.} \qquad I = J \end{cases} \qquad (8)$$

By adopting the AL method, the constrained problem (8) can be transformed into

$$\min_{I,J} \left\{ \sum_{p}^{q} \left[\psi_p J - \lambda_p (I - J) + \frac{\delta_p}{2} \| I - J \|_F^2 \right] + \frac{\zeta}{2} \left\| E - \frac{\gamma}{\zeta} - OI \right\|_F^2 \right\}, \qquad (9)$$

where $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_q\}$ and $\delta = \{\delta_1, \delta_2, \ldots, \delta_q\}$ are the Lagrangian multipliers and penalty parameters, respectively, which are associated with J in different transformations. By transformation, problem (9) can be further written as

$$\min_{I,J} \left\{ \sum_{p}^{q} \left(\psi_p J + \frac{\delta_p}{2} \left\| I - \frac{\lambda_p}{\delta_p} - J \right\|_F^2 \right) + \frac{\zeta}{2} \left\| E - \frac{\gamma}{\zeta} - OI \right\|_F^2 \right\}. \qquad (10)$$

Problem (10) can be solved by an alternating direction method of multipliers (ADMM) based on an iteration of solving the I-subproblem and J-subproblem alternatively. However, in the J-subproblem, the sparse domains in different transformations lead to different solutions at the beginning of the iteration. Therefore, some independent auxiliary variables $W = \{W_1, W_2, \ldots, W_q\}$ are introduced for each transformation, and thus problem (10) can be written as

$$\min_{I,W} \left\{ \sum_{p}^{q} \left(\psi_p W_p + \frac{\delta_p}{2} \left\| I - \frac{\lambda_p}{\delta_p} - W_p \right\|_F^2 \right) + \frac{\zeta}{2} \left\| E - \frac{\gamma}{\zeta} - OI \right\|_F^2 \right\}. \qquad (11)$$

To solve problem (11), the ADMM is also adopted to solve the I-subproblem and W-subproblem alternatively. In the kth iteration, the I-subproblem can be written as

$$I^k = \arg \min_{I} \left\{ \sum_{p}^{q} \frac{\delta_p^k}{2} \left\| I - \frac{\lambda_p^k}{\delta_p^k} - W_p^{k-1} \right\|_F^2 + \frac{\zeta^k}{2} \left\| E - \frac{\gamma^k}{\zeta^k} - OI \right\|_F^2 \right\}, \qquad (12)$$

and the W-subproblem can be written as

$$W_p^k = \arg \min_{W_p} \psi_p W_p + \frac{\delta_p^k}{2} \left\| I^k - \frac{\lambda_p^k}{\delta_p^k} - W_p \right\|_F^2. \qquad (13)$$

The I-subproblem in Eq. (12) is a quadratic regularized least-squares problem, and its direct solution is given in a closed form as

$$\begin{cases} I^k = \mathbf{HARD} \left[\zeta^k O^T E - \gamma^k O^T + \left(\sum_{p}^{f} \delta_p^k W_p^{k-1} - \lambda_p^k \right) \right], \\ \mathbf{HARD} = \left(\zeta^k O^T O + \sum_{p}^{f} \delta_p^k I_{\text{iden}} \right)^{-1}, \end{cases} \qquad (14)$$

where I_{iden} is an identity operator. It is expensive to compute **HARD** due to large data, so here the GD algorithm is used to solve the I-subproblem. The main shortcoming of the GD algorithm is that the number of iterations is very large because it is difficult to choose the step size. Some methods have been developed to obtain a better step size through much computation, such as the Barzilai and Borwein (BB) method [43,44]. Here, a learning method is used to seek the optimized step size. In our method, the number of iterations is much less than that of the BB method. Based on the GD algorithm, the solution to Eq. (12) can be expressed as

$$I^k = I^{k-1} - \alpha^k \left[\sum_{p}^{q} \delta_p^k \left(I^{k-1} - \frac{\lambda_p^k}{\delta_p^k} - W_p^{k-1} \right) - \zeta^k O^T \left(E - \frac{\gamma^k}{\zeta^k} - OI^{k-1} \right) \right], \qquad (15)$$

where α is the step size in the GD algorithm, and here the GD algorithm is utilized to calculate Eq. (15). From Eqs. (15) and (13), one can see that the solution to problem (6) depends mainly on W_p instead of I. For convenience, the solution to W_p can be further written as

$$W_p^k = S_p \left(I^k - \frac{\lambda_p^k}{\delta_p^k} \right). \qquad (16)$$

To obtain the solver $S_p(\cdot)$, the traditional algorithms usually employ the explicit handmade image prior as the sparse domain, such as a total variation (TV) prior and a wavelet prior [29,36]. However, the hand-crafted image prior has no pertinence for one dynamic scene, so it is not the best sparse domain. Here, we propose to learn the solver $S_p(\cdot)$ by convolutional neural networks. The architecture of learning solver $S_p(\cdot)$ is a spatial–temporal network, which is utilized to exploit the sparse domain from spatial and temporal correlation. This network consists of two sets of convolutional layers followed by a rectified linear unit (ReLU) layer and a single convolutional layer, as shown in Fig. 1(a), which is motivated by a recent work on image spatial super-resolution [45].

The general framework of the AL + DL algorithm is shown in Fig. 1(b). Compared with the conventional TwIST or AL algorithm, we optimize the sparse domain and some relevant iteration parameters $\{\alpha^k, \delta^k, \zeta^k\}$ by end-to-end training. The optimized sparse domain by specific training can greatly reduce the sparsity s and coherence μ, which is very helpful for high-fidelity image reconstruction, as shown in Eq. (4). To help denoise and retain more details after the iteration, we add U-net architecture containing self-attention, as shown in Fig. 2. The U-net has five times downsampling and upsampling, as shown in Fig. 2(a). In particular, we have two times convolution operations with stride 1 after downsampling or upsampling. Also, we impose self-attention to the layer that has
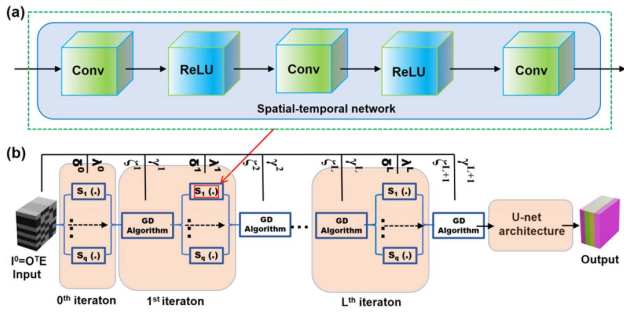
Fig. 1. Data flow chart of the AL + DL algorithm. (a) Solver $S_p(\cdot)$ in a sparse domain; (b) general framework by connecting each iteration in a sequence order. Here, each $S_i(\cdot)$ is calculated in parallel to producing W_i, and GD algorithm is employed to calculate I.

128 feature maps before deconvolution, which can help the architecture learn the long-range similarity easily, as shown in Fig. 2(b). Here, U-net allows the network to propagate the context information to some higher resolution layers, which has been successfully utilized to recover 3D information from 2D information in the spectral images [38]. Meanwhile, the self-attention mechanism, which has been recently proposed in computer vision tasks [46–49], can be used to exploit both the non-local similarity of spatial textures and the long-range temporal similarity, because the self-attention can help networks focus on some specific details and form some local specific feature. By embedding U-net architecture, the mean peak signal-to-noise ratio (PSNR) value of all the images in our simulating dynamic scenes increases by 0.81 dB, while the mean structural similarity index (SSIM) value increases by 0.007. Therefore, the AL + DL algorithm can retain more spatial details and finally achieve higher image fidelity than conventional AL and TwIST algorithms in theory. To facilitate researchers in citing and using our AL + DL algorithm, the codes are available at https://github.com/integritynoble/ALDL-algorithm.
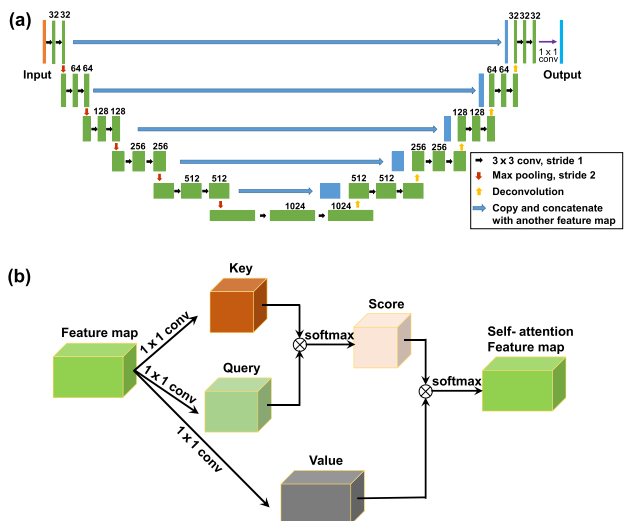

Fig. 2. (a) U-net architecture in the AL + DL algorithm; (b) self-attention model.

3. THEORETICAL SIMULATIONS

To validate the superior performance of the AL + DL algorithm in CUP, we perform three theoretical simulations and two experiments. In image reconstruction, TensorFlow is employed to implement the AL + DL algorithm on an NVIDIA Geforce GTX 2080Ti GPU with 11 GB device memory. Initially, the size of all images should be resized to $32N \times 32M$ due to five ($2^5 = 32$) times downsampling and upsampling in the U-net architecture, but the number K of frames is not limited, which indicates that the dynamic scene should have $32N \times 32M \times K$ cube, where N, M, and K can be adjusted according to the real dynamic scene. In fact, the resizing of the image has no side effect on the dynamic scene, because the size of images can be set to be larger than the actual one by padding zeros. When learning the model, the relevant iteration parameters are set as follows: all initial elements in Lagrangian multipliers γ and $\lambda = \{\lambda_1, \lambda_2, ..., \lambda_q\}$ are set to zero, initial I is set as $O^T E$, the number of iterations is 11, maximum running epoch is 280, and the initial learning rate is 0.008. Meanwhile, a rooted square-mean-error (RMSE) is used as the training loss, which is minimized by the Adam optimizer [50]. In each iteration, the values of Lagrangian multipliers γ and λ are calculated with the AL algorithm [24,28,29,51]. In our theoretical simulations, we chose three kinds of dynamic scenes with different complexities to test the ability of the AL + DL algorithm in the image reconstruction of CUP, and each dynamic scene contains eight frames. The three dynamic scenes are boatman [52], ocean animal [53], and finger [54]. Here, the boatman scene has some droplets and subtle textures, and therefore the relevant images are difficult to compress, representing the complex scene, while the finger scene contains only finger movement; thus, the relevant images are easy to compress, representing a simple scene. Usually, the inverse of the lossless compression ratio of images can be used to illustrate the complexity of a dynamic scene [55]. For each dynamic scene, 512 relevant pictures are utilized to train the model, and k-fold cross-validation is used to track the training effect. Here, this set of pictures is divided into two parts: one is used as training images, and the other is used as test images. To train the model, the 512 pictures are grouped and then combined into many small videos, and each video contains eight pictures, which corresponds to the frame number of each dynamic scene. Here, only one picture is replaced in each video compared to the previous video. Also, these original videos are randomly partitioned into eight equal-sized sub-videos in the eight-fold cross-validation. To show the superiority of the AL + DL algorithm, the AL and TwIST algorithms are also used for reconstruction based on the TV domain, which are used mostly for CUP [9–11,18–24, 56,57]. The reconstructed images of the boatman, ocean animal, and finger by the AL + DL, AL, and TwIST algorithms are shown in Fig. 3, together with the ground truth for comparison.

Here, only three representative pictures are selected, and an interesting area in each dynamic scene is enlarged for observation. Spatial details in the boatman, ocean animal, and finger can be clearly observed by the AL + DL algorithm, while these details are submerged by the AL and TwIST
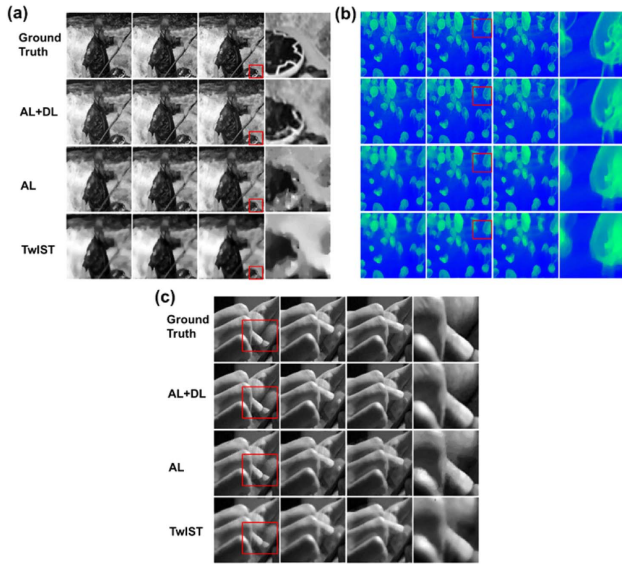
Fig. 3. Reconstructed results of (a) boatman, (b) ocean animal, and (c) finger by the AL + DL (second row), AL (third row), and TwIST (fourth row) algorithms, together with the ground truth (first row) for comparison. The last column is the enlarged image in the corresponding red squares.
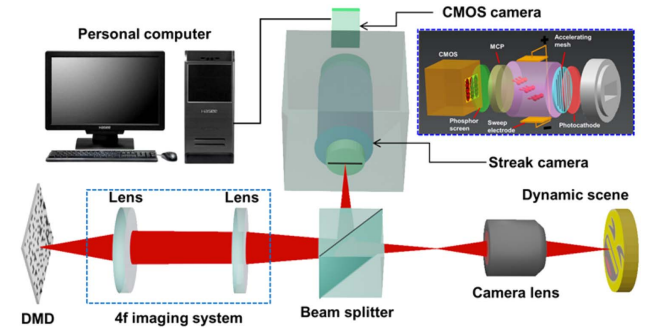

Fig. 4. System configuration of CUP. DMD, digital micromirror device; CMOS, complementary metal–oxide-semiconductor.

algorithms, which is disadvantageous for high-spatial-resolution imaging of a dynamic scene. To intuitively compare the improved efficiency in image fidelity by the AL + DL algorithm, we calculate PSNR and SSIM, and the calculated results are given in Table 1. Compared to the AL and TwIST algorithms, both PSNR and SSIM by the AL + DL algorithm are significantly improved. Here, PSNR (SSIM) is increased by at least 4.35 dB (0.136) for the boatman, 5.47 dB (0.114) for the ocean animal, and 9.78 dB (0.051) for the finger. Based on these calculated results, a rule can be found, which is, the simpler the spatial structure of the dynamic scene, the higher the improvement efficiency of PSNR, while the improvement efficiency of SSIM shows the opposite behavior. This phenomenon should be related to the sparsity of the dynamic scene; the simpler dynamic scene usually has higher sparsity, and vice versa. PSNR is based on a logarithmic function, which is not very well matched to perceived visual quality, but SSIM is based on visible structures in the image. Thus, PSNR has high improvement efficiency for a simple dynamic scene (i.e., finger), while SSIM has high improvement efficiency for a complex dynamic scene (i.e., boatman). In addition, the AL + DL algorithm can reconstruct a dynamic scene in only a few seconds,

which is much shorter than the AL and TwIST algorithms, which need tens of seconds; the computing efficiency is improved by an order of magnitude, which is very beneficial in practical applications of CUP.

4. EXPERIMENTAL RESULTS

Besides the above theoretical simulations, we also experimentally verify the superiority of the AL + DL algorithm on image reconstruction of CUP. The system configuration of CUP is given in Fig. 4. The dynamic scene is imaged via a camera lens and a $4f$ imaging system. On the image plane, a digital micromirror device (DMD) (Texas Instruments, DLP LightCrafter) is used to encode the dynamic scene in the spatial domain with a pseudo-random binary pattern, as encoding operator C. Through the collection of the same $4f$ imaging system and the reflection of a beam splitter, the encoded dynamic scene is vertically deflected by a streak camera (Hamamatsu, C7700), as shearing operator S. Finally, a complementary metal–oxide-semiconductor (CMOS) camera (Hamamatsu, ORCA-flash4.0) is employed to detect the encoded and deflected dynamic scene, as integrating operator T. Combining the measured image by CMOS and the codes on DMD, the original dynamic scene is reconstructed by the AL + DL, AL, and TwIST algorithms. For the training data of the AL + DL algorithm, we simulated the dynamic scenes based on the static images recorded without operators C and S.

First, we measure the temporal evolution of a spatially modulated picosecond laser spot, and the experimental design is shown in Fig. 5(a) [24]. The output 50 fs (full width at half maximum, FWHM) laser pulse from a Ti:sapphire amplifier is broadened to about 16 ps by a stretcher, and a thin wire is used to divide the laser spot into two components in space to obtain such a dynamic scene with special spatial structure. The spatially modulated laser spot illuminates a thin white paper, and a small fraction of photons can pass through the thin white paper. Thus, the temporal evolution behavior of a spatially modulated laser spot can be measured by our CUP system with a frame rate of 500 billion frames per second (fps). In this dynamic scene, the signal strength changes, while the spatial structure remains unchanged. The reconstructed images by the AL + DL, AL, and TwIST algorithms are shown in Figs. 5(b)–5(d), respectively. Compared with the AL and
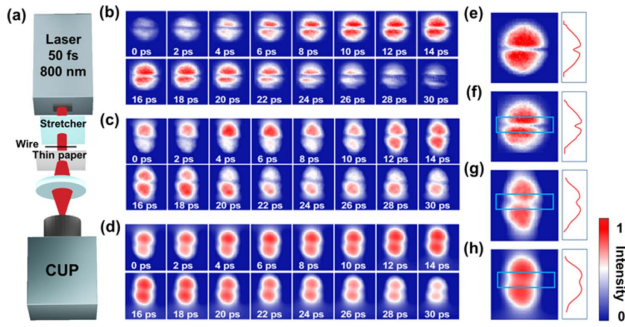
Table 1. Average PSNR (in dB) and SSIM by Different Image Reconstruction Algorithms in Different Dynamic Scenes

Scene	AL + DL		AL		TwIST	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Boatman	28.50	0.836	24.15	0.700	22.47	0.589
Ocean animal	30.47	0.916	25.00	0.802	24.72	0.781
Finger	42.00	0.983	32.22	0.932	28.56	0.894

Fig. 5. Measuring temporal evolution of a spatially modulated pico-second laser spot. (a) Experimental design. (b)–(d) Reconstructed results by the AL + DL, AL, and TwIST algorithms, respectively. (e) Measured static image by external CCD. (f)–(h) Extracted images from (b)–(d), respectively, at the time of 14 ps; curves on the right are the integration results of the corresponding images along the horizontal direction.

Fig. 6. Measuring wavefront movement by obliquely illuminating a collimated femtosecond laser pulse on a transverse fan pattern. (a) Experimental design. (b)–(d) Reconstructed results by the AL + DL, AL, and TwIST algorithms, respectively. (e) Measured static image by external CCD. (f)–(h) Integrated images from (b)–(d), respectively. (i)–(l) Results of Fourier transform from (e)–(h), respectively.

TwIST algorithms, the reconstructed images by the AL + DL algorithm have a clearer spatial shape and less background noise. To further compare the image fidelity by the three algorithms, we chose the reconstructed images at a time of 14 ps to compare with the static image, as shown in Figs. 5(e)–5(h).

Here, the static image is achieved by external CCD measurement without encoding operator C and shearing operator S, as shown in Fig. 5(e). Meanwhile, the intensities of Figs. 5(e)–5(h) are also integrated along the horizontal direction, and the calculated results are given on the right of the relative images. The AL + DL algorithm can retain very high image fidelity, but the AL and TwIST algorithms cause a certain degree of image distortion. The fundamental reason should be the mismatch of the sparse domain in image reconstruction. More importantly, like the static image, the blocked part in the laser spot (see light blue squares) can be clearly distinguished by the AL + DL algorithm, where an obvious valley in the intensity curve is observed, but not by either the AL or TwIST algorithm, especially the TwIST algorithm.

In the first experiment in Fig. 5(a), the spatial shape of the dynamic scene remains unchanged. In the second experiment, we measure the wavefront movement by obliquely illuminating a collimated femtosecond laser pulse on a transverse fan pattern, where both the signal strength and spatial shape in the dynamic scene change. The experimental design is presented in Fig. 6(a). A 7 ps (FWHM) laser pulse after collimation obliquely illuminates a transverse fan pattern with an angle of ∼30° to the surface normal. Our CUP system faces the pattern surface and collects the scattered photons from the pattern scene. Here, the shearing velocity of the streak camera is 0.66 km/s; thus, the imaging speed is 50 billion fps, i.e., 20 ps exposure time in theory [9]. The reconstruction images by the AL + DL, AL, and TwIST algorithms are presented in Figs. 6(b)–6(d), respectively. As expected, the spatial shape of the fan can be displayed in the whole process of wavefront movement by the AL + DL algorithm for image reconstruction, while it is blurred by the AL and TwIST algorithms due to the artifacts in the image reconstruction. To better evaluate the image reconstruction effect of the three
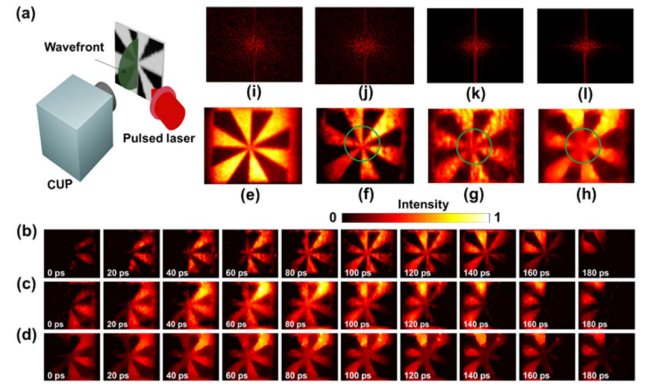
algorithms, the reconstructed images in Figs. 6(b)–6(d) are integrated and compared to the static image measured by an external CCD, as shown in Figs. 6(e)–6(h). Similar to the static image, the whole outline of the fan in the integrated image via the AL + DL algorithm is clear, but it is a little fuzzy by the AL and TwIST algorithms, especially for the center part of the fan (green circles). To intuitively illustrate the spatial resolution, the images in Figs. 6(e)–6(h) are processed via Fourier transform, and the calculated results are shown in Figs. 6(i)–6(l). As can be seen, the AL + DL algorithm can obtain high-frequency information, which is almost the same as the static image, while the high-frequency information is lost for the AL and TwIST algorithms. In general, high-frequency information represents the fine structure in the spatial domain. Therefore, compared to the AL and TwIST algorithms, the AL + DL algorithm has great advantages in observing the spatial details of a complex dynamic scene.

5. DISCUSSION

The AL + DL algorithm is a data-driven method, which can optimize the sparse domain and relevant iteration parameters by learning instead of hand-crafted determination. For CS, the sparse domain is the core part that determines the sparsity and affects mainly the coherence. Thus, the sparse domain almost determines the image reconstruction quality. In general, the learning method can seek better sparse domain and iteration parameters, and therefore the AL + DL algorithm can get higher image fidelity than conventional AL and TwIST algorithms. Because of learning the sparse domain and iteration parameters, the AL + DL algorithm has high robustness and allows the encoding operator C to be different in training and testing processes, while the pure neural network algorithms cannot, such as deep fully connected networks [58], ReconNet [59], DR²-Net [60], λ-net [38], and DeepCubeNet [61]. Also, the AL + DL algorithm embeds a GD algorithm into tensor

computation, which involves massive data. In calculation, the GD algorithm does not easily find the appropriate step size, so it needs to perform many iterations, i.e., the convergence speed is low. To decrease the number of iterations, data scientists prefer Newton's method or a conjugate gradient algorithm by increasing the cost of each iteration [62]. However, some mathematicians seek a better step size in the GD algorithm to decrease the number of iterations by increasing the cost of each iteration, such as the BB method. Here, we utilize the GD algorithm to calculate the large data by a data-driven method based on the learning model, which can find the optimal step size to decrease the number of iterations without increasing the cost of each iteration and make the gradient show better orthogonality. It is noted that the AL + DL algorithm needs just 15 iterations, while the corresponding traditional algorithm based on the BB method needs more than 100 iterations.

As shown in Figs. 3, 5, and 6, compared to the AL and TwIST algorithms, the AL + DL algorithm shows great advantages in image reconstruction accuracy, but it also inherits the shortcoming of the data-driven method, i.e., the dependence on a learning dataset. In image reconstruction, these images in the dataset should have some similarities to those in the dynamic scene. An inappropriate training dataset may lead to results worse than those obtained by the AL and TwIST algorithms. In some special dynamic scenes, it may be difficult to find a similar dataset for training. In this case, it is feasible to increase the sampling rate m, such as lossless-CUP or multi-encoding CUP. Moreover, it is also a good idea to optimize the codes, which is similar to optimizing the sparse domain, which can reduce coherence. However, the AL + DL algorithm cannot be adopted directly to optimize the codes, because here the codes are considered as constant. Optimizing the codes demands that the mathematical architecture regard the codes as a variable; thus, the whole AL + DL architecture needs to be redesigned. In the future, we will strive to seek some new algorithms to simultaneously optimize the codes, sparse domain, and iteration parameters by learning the dataset.

6. CONCLUSION

In summary, we have developed a new AL + DL algorithm to realize high-fidelity image reconstruction for CUP. In our method, there are four key points: (1) optimizing the sparse domain in multiple transformation; (2) optimizing the relevant calculation parameters in the iteration process; (3) employing the GD algorithm to improve computing efficiency; (4) embedding the U-net architecture to help denoise. Key points (1), (2), and (4) are implemented by the DL method, and improving key point (3) also needs the DL method. However, the whole framework is determined by the AL method, which combines these four key points. Thus, the AL + DL algorithm not only utilizes the training neural networks, but also has some potential mathematical interpretations. More importantly, these results from theoretical simulations and experimental measurements show that the AL + DL algorithm is superior to conventional AL and TwIST algorithms in image fidelity and computing efficiency. Additionally, the AL + DL algorithm is a simple mathematical architecture, so it is easy to extend to other high-dimensional tensor fields. In future studies, we will continue to search for better image reconstruction algorithms for CUP to achieve super-high image fidelity.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. P. R. Poulin and K. A. Nelson, "Irreversible organic crystalline chemistry monitored in real time," Science **313**, 1756–1760 (2006).
2. P. Hockett, C. Z. Bisgaard, O. J. Clarkin, and A. Stolow, "Time-resolved imaging of purely valence-electron dynamics during a chemical reaction," Nat. Phys. **7**, 612–615 (2011).
3. R. Horstmeyer, H. Ruan, and C. Yang, "Guidestar-assisted wavefront-shaping methods for focusing light into biological tissue," Nat. Photonics **9**, 563–571 (2015).
4. J. W. Borst and A. J. Visser, "Fluorescence lifetime imaging microscopy in life sciences," Meas. Sci. Technol. **21**, 102002 (2010).
5. H. R. Petty, "Spatiotemporal chemical dynamics in living cells: from information trafficking to cell physiology," Biosystems **83**, 217–224 (2006).
6. T. M. Squires and S. R. Quake, "Microfluidics: fluid physics at the nanoliter scale," Rev. Mod. Phys. **77**, 977–1026 (2005).
7. N. Šiaulys, L. Gallais, and A. Melninkaitis, "Direct holographic imaging of ultrafast laser damage process in thin films," Opt. Lett. **39**, 2164–2167 (2014).
8. R. L. Kodama, P. A. Norreys, K. Mima, A. E. Dangor, R. G. Evans, H. Fujita, Y. Kitagawa, K. Krushelnick, T. Miyakoshi, and N. Miyanaga, "Fast heating of ultrahigh-density plasma as a step towards laser fusion ignition," Nature **412**, 798–802 (2001).
9. L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**, 74–77 (2014).
10. J. Liang, L. Zhu, and L. V. Wang, "Single-shot real-time femtosecond imaging of temporal focusing," Light Sci. Appl. **7**, 42 (2018).
11. D. Qi, S. Zhang, C. Yang, Y. He, F. Cao, J. Yao, P. Ding, L. Gao, T. Jia, and J. Liang, "Single-shot compressed ultrafast photography: a review," Adv. Photon. **2**, 014003 (2020).
12. K. Nakagawa, A. Iwasaki, Y. Oishi, R. Horisaki, A. Tsukamoto, A. Nakamura, K. Hirosawa, H. Liao, T. Ushida, and K. Goda, "Sequentially timed all-optical mapping photography (STAMP)," Nat. Photonics **8**, 695–700 (2014).
13. T. Suzuki, R. Hida, Y. Yamaguchi, K. Nakagawa, T. Saiki, and F. Kannari, "Single-shot 25-frame burst imaging of ultrafast phase transition of $Ge_2Sb_2Te_5$ with a sub-picosecond resolution," Appl. Phys. Express **10**, 092502 (2017).
14. Y. Lu, T. T. Wong, F. Chen, and L. Wang, "Compressed ultrafast spectral-temporal photography," Phys. Rev. Lett. **122**, 193904 (2019).
15. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**, 745 (2012).
16. A. H. Zewail, "Four-dimensional electron microscopy," Science **328**, 187–193 (2010).
17. A. Barty, S. Boutet, M. J. Bogan, S. Hau-Riege, S. Marchesini, K. Sokolowski-Tinten, N. Stojanovic, H. Ehrke, A. Cavalleri, and S. Düsterer, "Ultrafast single-shot diffraction imaging of nanoscale dynamics," Nat. Photonics **2**, 415–419 (2008).
18. J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, "Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse," Sci. Adv. **3**, e1601814 (2017).

19. J. C. Jing, X. Wei, and L. V. Wang, "Spatio-temporal-spectral imaging of non-repeatable dissipative soliton dynamics," Nat. Commun. **11**, 2059 (2020).

20. T. Kim, J. Liang, L. Zhu, and L. V. Wang, "Picosecond-resolution phase-sensitive imaging of transparent objects in a single shot," Sci. Adv. **6**, e6200 (2020).

21. J. Liang, L. Gao, P. Hai, C. Li, and L. V. Wang, "Encrypted three-dimensional dynamic imaging using snapshot time-of-flight compressed ultrafast photography," Sci. Rep. **5**, 15504 (2015).

22. F. Cao, C. Yang, D. Qi, J. Yao, Y. He, X. Wang, W. Wen, J. Tian, T. Jia, and Z. Sun, "Single-shot spatiotemporal intensity measurement of picosecond laser pulses with compressed ultrafast photography," Opt. Laser Eng. **116**, 89–93 (2019).

23. L. Zhu, Y. Chen, J. Liang, Q. Xu, L. Gao, C. Ma, and L. V. Wang, "Space-and intensity-constrained reconstruction for compressed ultrafast photography," Optica **3**, 694–697 (2016).

24. C. Yang, D. Qi, F. Cao, Y. He, X. Wang, W. Wen, J. Tian, T. Jia, Z. Sun, and S. Zhang, "Improving the image reconstruction quality of compressed ultrafast photography via an augmented Lagrangian algorithm," J. Opt. **21**, 035703 (2019).

25. Y. Lai, Y. Xue, C. Y. Côté, X. Liu, A. Laramée, N. Jaouen, F. Légaré, L. Tian, and J. Liang, "Single-shot ultraviolet compressed ultrafast photography," Laser Photon. Rev. **14**, 2000122 (2020).

26. C. Yang, D. Qi, X. Wang, F. Cao, Y. He, W. Wen, T. Jia, J. Tian, Z. Sun, and L. Gao, "Optimizing codes for compressed ultrafast photography by the genetic algorithm," Optica **5**, 147–151 (2018).

27. C. Yang, D. Qi, J. Liang, X. Wang, F. Cao, Y. He, X. Ouyang, B. Zhu, W. Wen, and T. Jia, "Compressed ultrafast photography by multi-encoding imaging," Laser Phys. Lett. **15**, 116202 (2018).

28. C. Li, "An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing," Master dissertation (Rice University, 2010).

29. M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," IEEE Trans. Image Process. **20**, 681–695 (2010).

30. Y. Yang, J. Sun, H. Li, and Z. Xu, "ADMM-CSNet: a deep learning approach for image compressive sensing," IEEE Trans. Pattern Anal. **42**, 521–538 (2018).

31. J. Zhang and B. Ghanem, "ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1828–1837.

32. J. Ma, X. Liu, Z. Shou, and X. Yuan, "Deep tensor ADMM-net for snapshot compressive imaging," in *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 10223–10232.

33. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. Express **27**, 28075–28090 (2019).

34. Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," IEEE Trans. Pattern Anal. Mach. Intell. **40**, 1888–1902 (2017).

35. Y. Wang, J. Peng, Q. Zhao, Y. Leung, X. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," IEEE J. Sel. Top. Appl. Earth Observ. Remote Sensing **11**, 1227–1243 (2017).

36. L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 8032–8041.

37. Z. Wu, Y. Sun, A. Matlock, J. Liu, L. Tian, and U. S. Kamilov, "SIMBA: scalable inversion in optical tomography using deep denoising priors," IEEE J. Sel. Top. Appl. Earth Observ. Remote Sensing **14**, 1163–1175 (2020).

38. X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "Lambda-net: reconstruct hyperspectral images from a snapshot measurement," in *IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2019), pp. 4058–4068.

39. J. M. Bioucas-Dias and M. A. Figueiredo, "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).

40. E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" IEEE Trans. Inform. Theory **52**, 5406–5425 (2006).

41. E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Commun. Pure Appl. Math. **59**, 1207–1223 (2006).

42. X. Liu and X. Wang, "Fourth-order tensors with multidimensional discrete transforms," arXiv:1705.01576 (2017).

43. J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," IMA J. Numer. Anal. **8**, 141–148 (1988).

44. M. Raydan, "Convergence properties of the Barzilai and Borwein gradient method," Ph.D. dissertation (Rice University, 1991).

45. B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 136–144.

46. L. Yue, X. Miao, P. Wang, B. Zhang, X. Zhen, and X. Cao, "Attentional alignment networks," in *29th British Machine Vision Conference* (2018), pp. 1–14.

47. S. Min, X. Chen, Z. Zha, F. Wu, and Y. Zhang, "A two-stream mutual attention network for semi-supervised biomedical segmentation with noisy labels," in *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), pp. 4578–4585.

48. Y. Li, Z. Xiao, X. Zhen, and X. Cao, "Attentional information fusion networks for cross-scene power line detection," IEEE Geosci. Remote Sens. Lett. **16**, 1635–1639 (2019).

49. Y. Huang, X. Cao, X. Zhen, and J. Han, "Attentive temporal pyramid network for dynamic scene classification," in *Proceedings of the AAAI Conference on Artificial Intelligence* (2019), pp. 8497–8504.

50. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).

51. S. H. Chan, R. Khoshabeh, K. B. Gibson, P. E. Gill, and T. Q. Nguyen, "An augmented Lagrangian method for video restoration," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2011), pp. 941–944.

52. https://drive.google.com/file/d/1PGav_1Yju2lfXlXQsxnvkMPu_j_EWm0B/view.

53. https://www.pexels.com/video/blue-jellyfish-856882/.

54. https://www.pexels.com/video/a-person-using-a-smart-cellular-phone-3143531/.

55. H. Yu and S. Winkler, "Image complexity and spatial information," in *Fifth International Workshop on Quality of Multimedia Experience (QoMEX)* (IEEE, 2013), pp. 12–17.

56. C. Yang, D. Qi, F. Cao, Y. He, J. Yao, P. Ding, X. Ouyang, Y. Yu, T. Jia, and S. Xu, "Single-shot receive-only ultrafast electro-optical deflection imaging," Phys. Rev. Appl. **13**, 024001 (2020).

57. C. Yang, F. Cao, D. Qi, Y. He, P. Ding, J. Yao, T. Jia, Z. Sun, and S. Zhang, "Hyperspectrally compressed ultrafast photography," Phys. Rev. Lett. **124**, 023902 (2020).

58. M. Iliadis, L. Spinoulas, and A. K. Katsaggelos, "Deep fully-connected networks for video compressive sensing," Digit. Signal Process. **72**, 9–18 (2018).

59. K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "ReconNet: non-iterative reconstruction of images from compressively sensed measurements," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 449–458.

60. H. Yao, F. Dai, S. Zhang, Y. Zhang, Q. Tian, and C. Xu, "DR^2-Net: deep residual reconstruction network for image compressive sensing," Neurocomputing **359**, 483–493 (2019).

61. D. Gedalin, Y. Oiknine, and A. Stern, "DeepCubeNet: reconstruction of spectrally compressive sensed hyperspectral images with deep neural networks," Opt. Express **27**, 35811–35822 (2019).

62. J. Nocedal and S. Wright, *Numerical Optimization* (Springer, 2006).