Research Article | Vol. 29, No. 10 / 10 May 2021 / Optics Express 15239
Optics EXPRESS

Two-step training deep learning framework for computational imaging without physics priors

RUIBO SHANG,[1] KEVIN HOFFER-HAWLIK,[1] FEI WANG,[2,3] GUOHAI SITU,[2,3,4] AND GEOFFREY P. LUKE[1,*]

[1]Thayer School of Engineering, Dartmouth College, 14 Engineering Dr., Hanover, NH 03755, USA
[2]Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China
[3]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[4]Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
*geoffrey.p.luke@dartmouth.edu

Abstract: Deep learning (DL) is a powerful tool in computational imaging for many applications. A common strategy is to use a preprocessor to reconstruct a preliminary image as the input to a neural network to achieve an optimized image. Usually, the preprocessor incorporates knowledge of the physics priors in the imaging model. One outstanding challenge, however, is errors that arise from imperfections in the assumed model. Model mismatches degrade the quality of the preliminary image and therefore affect the DL predictions. Another main challenge is that many imaging inverse problems are ill-posed and the networks are over-parameterized; DL networks have flexibility to extract features from the data that are not directly related to the imaging model. This can lead to suboptimal training and poorer image reconstruction results. To solve these challenges, a two-step training DL (TST-DL) framework is proposed for computational imaging without physics priors. First, a single fully-connected layer (FCL) is trained to directly learn the inverse model with the raw measurement data as the inputs and the images as the outputs. Then, this pre-trained FCL is fixed and concatenated with an un-trained deep convolutional network with a U-Net architecture for a second-step training to optimize the output image. This approach has the advantage that does not rely on an accurate representation of the imaging physics since the first-step training directly learns the inverse model. Furthermore, the TST-DL approach mitigates network over-parameterization by separately training the FCL and U-Net. We demonstrate this framework using a linear single-pixel camera imaging model. The results are quantitatively compared with those from other frameworks. The TST-DL approach is shown to perform comparable to approaches which incorporate perfect knowledge of the imaging model, to be robust to noise and model ill-posedness, and to be more robust to model mismatch than approaches which incorporate imperfect knowledge of the imaging model. Furthermore, TST-DL yields better results than end-to-end training while suffering from less overfitting. Overall, this TST-DL framework is a flexible approach for image reconstruction without physics priors, applicable to diverse computational imaging systems.

1. Introduction

Computational Imaging is a powerful tool in the application of image reconstruction. It relaxes the hardware requirements of imaging systems by relying on (typically iterative) computational techniques to recover the lost information, that is, solving an inverse imaging problem computationally [1,2]. These methods rely on a measured or assumed forward operator of the imaging system to create a mapping from the image to the measurement. However, the inverse problem is often ill-posed by design or due to the imperfect physical measurement, meaning multiple

#424165 https://doi.org/10.1364/OE.424165
Journal © 2021 Received 2 Mar 2021; revised 21 Apr 2021; accepted 23 Apr 2021; published 3 May 2021

Optics EXPRESS

solutions exist for a given measurement. Therefore, additional information about the scene or the object must be incorporated in the computational process for accurate reconstruction.

One of the most common methods in computational imaging is sparsity-based optimization which seeks to reconstruct images from incomplete measurements or an ill-posed inverse problem [3,4]. This concept is based on the knowledge that most natural images are sparse (i.e., only a few nonzero values exist) when transformed into a specific domain. Researchers have successfully applied sparsity-based optimization in a variety of imaging fields ranging from compressed ultrafast photography [5] to holographic video [6] to biomedical imaging [7]. Although sparsity-based optimization has advantages in image reconstruction, the primary drawback to this approach is that it is iterative and time consuming. Depending on the scale and scope of the problem, an image reconstruction task can require minutes to even hours of computation. Therefore, it cannot achieve real-time imaging for many applications which require pipelined data acquisition and image reconstruction. Furthermore, the optimal algorithm-specific parameters in the sparsity-based optimization framework must be heuristically determined.

Deep learning (DL) [2,8] is an emerging computational imaging approach dramatically improving the state-of-the-art in fast image reconstruction [9–15]. Instead of building a specific model and finding the optimal algorithm-specific parameters heuristically (as in sparsity-based optimization approaches), it relies on large amounts of data to automatically learn tasks by finding the optimal parameters in each layer of a neural network [8]. It has the benefit of being computationally efficient since most of the computational energy is used during the one-time training process. Compared with sparsity-based optimization approaches which require iterative testing of the regularizer for each image [16], DL approaches utilize the training dataset to find the optimal regularizer for a broad range of images. Therefore, DL is a promising alternative to augment or replace the iterative algorithms used in sparsity-based optimization. Researchers have applied the DL approach in many imaging fields with varying network structures [2]. The U-Net [17] architecture is one of the most successful DL frameworks in the imaging field. Its architecture consists of a contracting path to capture context and a symmetric expanding path for enhancement of key features. Skip connections between the contracting and expanding path help to preserve features from the input image. A variety of applications in the imaging field, ranging from segmentation to image reconstruction from incomplete data, have harnessed the original or a modification of the U-Net structure [18–24].

In many cases, the acquired data can be directly fed into an end-to-end DL framework to reconstruct an image [9,13]. A preprocessing step, however, can ease the burden on the network by forming an initial estimate of the image with the knowledge of physics priors [2]. This is particularly helpful when the data are noisy [14], when the acquired data have a different size or dimensionality than the reconstructed image [25], or when the data are acquired in a different domain than the image (e.g., the Fourier domain) [22,26,27]. The preprocessing step typically takes the form of a computational image reconstruction algorithm which incorporates an approximation of the imaging forward model [14,22,28,29]. This step can be computationally intensive, especially when using iterative image reconstruction approaches [30,31]. Furthermore, the forward model in many imaging fields can be difficult to acquire with high accuracy (i.e., a model mismatch exists) [7,32,33]. The model mismatch will lead to an inaccurate initial image guess and therefore affects the DL prediction. Furthermore, deep learning networks are over-parameterized, meaning they have the potential to adapt to a wide range of data types and imaging problems. This also means, however, that the networks have the flexibility to extract features from the data that are not directly related to the imaging model (there are insufficient constraints to enforce learning of the physical model rather than extraction of image features) [34]. This makes the networks susceptible to the model ill-posedness, leading to a degradation in performance when they encounter new data.

Optics EXPRESS

In this paper, a two-step training DL (TST-DL) framework is proposed for DL-based computational image reconstruction without prior knowledge of the model. The first (preprocessing) step trains a single fully-connected layer (FCL) to approximate the imaging inverse model. The weights of this trained FCL are then fixed and concatenated with an untrained convolutional neural network (U-Net) for second-step training to effectively impose regularization constraints and improve the reconstruction quality of the results predicted from the first-step training. The TST-DL approach can be applied to diverse computational imaging problems which require a preprocessing step because it places no constraints on the input size or dimensionality. We demonstrate that the method is robust to noisy data and ill-posed problems while yielding comparable results to methods incorporating ideal physics priors. The results show that by splitting the training process into two steps, TST-DL is resistant to the effects of over-parameterization observed in end-to-end DL approaches. Finally, the approach does not rely on prior knowledge of the imaging model. Thus, errors arising from an incorrect or uncertain model are avoided. Overall, these results also provide insight into how noise and model mismatch may affect the decision of when and how to apply physics priors in computational imaging problems.

2. Methods

2.1. Regularized optimization

Any linear imaging model can be described by

$$g = \boldsymbol{H}f + n \tag{1}$$

where f is the image to be reconstructed, g is the raw measurement, n is the noise and $\boldsymbol{H}$ is the forward operator.

The most straightforward way to reconstruct the image f is to find the inverse of the forward operator $\boldsymbol{H}^{-1}$ so that $\boldsymbol{H}^{-1}\boldsymbol{H} = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix. However, for most of the cases, $\boldsymbol{H}^{-1}$ is not unique or requires excessive computational power to determine.

An effective alternative to directly computing the inverse of the forward model is to iteratively solve the optimization problem,

$$\hat{f} = argmin_f ||\boldsymbol{H}f - g||_2^2 \tag{2}$$

where $||\cdot||_2$ denotes the L_2 norm. However, this pseudo-inverse solution is prone to artifacts and noise due to the ill-posed property of the corresponding inverse problem. Therefore, additional information is needed to converge to the correct solution.

A regularized optimization approach can incorporate additional knowledge about the image by adding a regularization term,

$$\hat{f} = argmin_f \{||\boldsymbol{H}f - g||_2^2 + \lambda\phi(f)\} \tag{3}$$

where ϕ is the regularization operator and λ is the regularization parameter. $||\boldsymbol{H}f - g||_2^2$ is the fidelity term and $\phi(f)$ is the regularization term. The regularization term is to make a balance with the fidelity term by driving the optimized $\hat{f}$ to match a specific regularization rule. Common regularization domains include spatial, edge, and wavelet domains. However, finding the optimal regularization rule for a specific image dataset is still a challenging problem [16].

Inspired by the regularized optimization approach, we propose the TST-DL framework. The first-step training is to train an FCL to learn an optimal linear $\boldsymbol{H}^{-1}$ given the training datasets (g,f). Then, this pre-trained FCL is fixed and concatenated with a U-Net for the second-step training to learn an optimal regularization rule to regularize f towards the optimal solution. By decoupling the whole inverse problem into two sub-problems with the two-step training strategy, we enforce the first step (FCL) to learn the model and the second step to learn the optimal regularizer (Section 3.2 for details).

Optics EXPRESS

2.2. TST-DL architecture

Our TST-DL framework contains an FCL and a U-Net architecture as shown in Fig. 1.

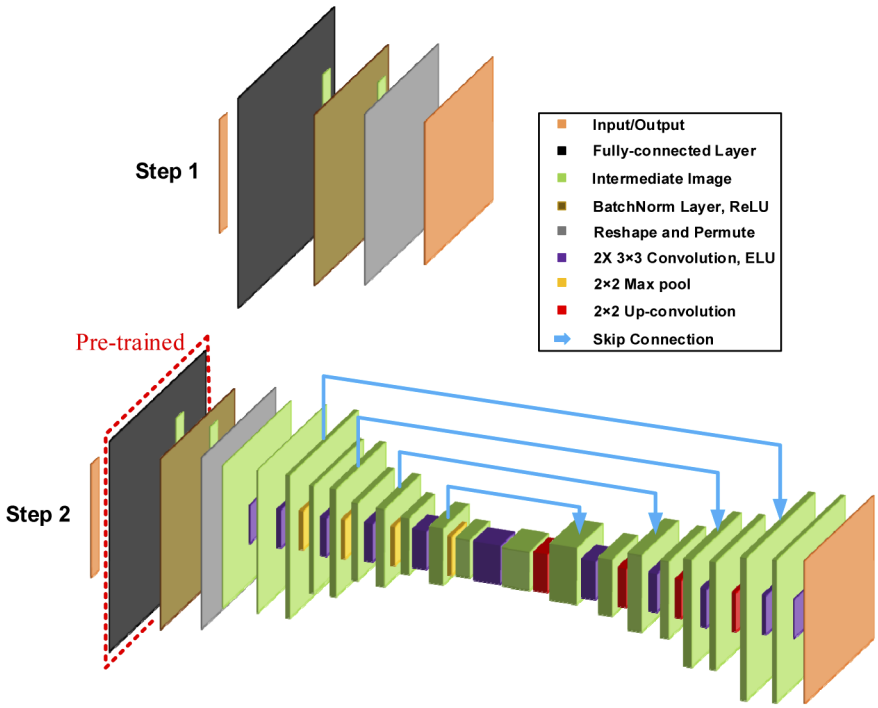

Fig. 1. TST-DL structure. Step 1 is training the FCL and step 2 is training a U-Net concatenated with the fixed pre-trained FCL. The input is the raw measurement data that can be any size and dimension and the output is a two-dimensional (2D) image.

The DL framework in step 1 consists of an FCL mapping from the raw measurement data (input) to the image (output) (Batch-normalization (BatchNorm), reshape and permute layers are used for normalization and reshape purposes). With this FCL, the input measurement data and the output image do not need to have the same size or even the same dimensionality. By training the FCL, the optimal inverse operator will be learned given the training datasets. The DL framework in step 2 follows the U-Net architecture [17] concatenated with the pre-trained FCL from step 1. The U-Net, which utilizes an encoder-decoder structure with skip connections to preserve wide-frequency features, was chosen because of its success in solving image-to-image problems. Dropout layers are included in each stage of the U-Net. The mean squared error (MSE) is used as the loss function in the first-step training to find the optimal H^{-1} that minimizes $||f - \boldsymbol{H}^{-1}g||_2^2$. A customized loss function with a combination of the root mean squared error (RMSE) and the difference of the structural similarity index (DSSIM) is used for the second-step training. The Adam optimizer is used with the default learning rate of 0.001. The batch size is chosen to be 50 and each training step runs 100 epochs.

2.3. Comparisons with other approaches

Quantitative comparisons are made with other DL frameworks (a deep convolutional auto-encoder network (DCAN) [35], two-step DCAN, one-step training DL (OST-DL) and the physics-prior-based DL (PPB-DL) approach with the U-Net architecture) and the established model-based optimization approaches (an iterative L_2 norm minimization approach LSQR [30] and a two-step

Optics EXPRESS

iterative shrinkage/thresholding (TwIST) algorithm [31]). The comparisons with DCAN, two-step DCAN, LSQR and TwIST are included in Supplement 1, Section 1. The DCAN is developed in single-pixel imaging to reconstruct the dynamic scenes from the single-pixel camera capture of the compressed signal. DCAN is comprised of two parts, the encoding part to find the optimal binary filters for the measurement and the decoding part for image reconstruction with FCL and three convolutional layers [35]. We only use the decoding part in DCAN since the binary filters as the physics priors are unknown. The two-step DCAN splits the training of the FCL and convolutional layers into two steps, similar to TST-DL. For the PPB-DL approach, an initial guess of the image is reconstructed using the LSQR approach. Then, the initial guess of the image is used as the input of U-Net for further training and prediction. For OST-DL, as an end-to-end DL approach, the FCL is concatenated with U-Net for single-step training to learn the inverse model and the optimal regularizer simultaneously instead of training each individually.

2.4. Imaging models and data acquisition

2.4.1. Model description and data simulation for single-pixel imaging

Single-pixel imaging [36, 37] with Russian-Doll (RD) Hadamard [38] patterns is used as an example of a linear imaging model. In RD Hadamard patterns, the measurement order of the Hadamard basis is reordered and optimized according to their significance for general scenes, such that at discretized increments, the complete sampling for different spatial frequencies is obtained [38]. The STL-10 natural image database [39] was used for training the TST-DL framework with 10,000 images as the training dataset, 2,000 images as the validating dataset and another 2,000 images as the testing dataset. In order to meet the dimension requirement of the RD Hadamard patterns, all the images were down-sampled from 96×96 to 64×64. The full RD Hadamard basis for a 64×64 image has 4,096 RD Hadamard patterns each with a size of 64×64. Different compression ratios (different levels of model ill-posedness) were used here as 2X, 4X and 16X corresponding to taking the first 1/2, 1/4 and 1/16 of RD Hadamard patterns, respectively. For instance, in 4X compression, the first 1,024 RD Hadamard patterns were used. The one-dimensional (1D) raw measurement data were acquired by multiplying each individual image with the RD Hadamard patterns at each compression ratio. Therefore, the 1D raw measurement data have a size of 2,048×1, 1,024×1 and 256×1 for the corresponding compression ratios. Different levels of white Gaussian noise (-5 dB, 0 dB and 10 dB signal-to-noise ratio (SNR) levels) were added to the 1D measurement data. For OST-DL as the one-step training approach, the training runs 200 epochs. For PPB-DL, since the initial guess of the image is obtained because of the known forward model, the training runs 100 epochs for a fair comparison. For TST-DL, each training step runs for 100 epochs.

2.4.2. Model description and experimental data acquisition for single-pixel imaging

Single-pixel imaging with random grayscale illumination patterns were conducted in the experiment. The images were taken from MNIST database [40] and resized from 28×28 to 32×32 pixels. 1,024 random grayscale illumination patterns each with a size of 32×32 were prepared as the full measurement basis. Then, the first 64 or 4 illumination patterns in the full basis were used to illuminate the objects, corresponding to a 16X or 256X compression ratio, respectively. Therefore, the corresponding 1D raw measurement data have a size of 64×1 or 4×1. The models were trained on an experimentally acquired dataset of 800 images and tested on 100 images. The batch size was chosen to be 40 since only 800 images were used for training. For TST-DL, each step was trained with 500 epochs. For fair comparison, PPB-DL was trained with 500 epochs since the prior knowledge of the model was provided and OST-DL was trained with 1,000 epochs. The imaging system is shown in Fig. 3 in [33]. Each image was displayed on a spatial light modulator (Pluto-Vis, Holoeye Photonics AG) and illuminated by a set of random

Optics EXPRESS

grayscale patterns from a digital micromirror device. The 1D measurement data were recorded by a bucket detector.

3. Simulation and analysis of TST-DL in single-pixel imaging

In this section, the single-pixel imaging [36,37] with RD Hadamard [38] patterns is used as the case of the linear imaging model to analyze TST-DL. The model description and the data acquisition are detailed in Section 2.4.1. The noise-free case (analyzed in Supplement 1, Section 1) showed the superior performance of U-Net-based architecture (TST-DL, OST-DL and PPB-DL). Thus, the remainder of the paper focuses on the U-Net architecture to more deeply investigate the two-step training process. The comparison between TST-DL and the DL approach with the physics priors (PPB-DL) in terms of model ill-posedness and noise, and the capability and robustness of the FCL to learn the inverse model are analyzed in Section 3.1. The advantage of the two-step training in TST-DL over the one-step training approach is analyzed in Section 3.2. Model mismatches are analyzed in Section 3.3 in detail to show under what circumstances a neural-network-based preprocessor should be used rather than the physics-prior-based preprocessor.

3.1. Robustness of TST-DL compared with the PPB-DL

Since the model ill-posedness and noise existing in measurement data are two major challenges in imaging inverse problems, we sought to explore if TST-DL has similar robustness to model ill-posedness and noise compared with a DL approach with physics priors (PPB-DL) in the predictions of inverse models. The mean and standard deviation of the RMSE and structural similarity index (SSIM) [41] for all the reconstructed images in the testing dataset from TST-DL and PPB-DL at each noise level and each compression ratio were calculated to quantitatively compare the performance. Figures 2(a) and 2(b) shows the quantitative results of the step 1 predictions of TST-DL (FCL of TST-DL) compared with the initial guess images of PPB-DL (Inputs of PPB-DL) from the LSQR approach. The reason that the RMSE and SSIM of the inputs of PPB-DL are worse than those of the FCL of TST-DL is that the FCL of TST-DL is trained on noisy data and thus incorporates denoising into its model inversion. Figures 2(c) and 2(d) show the final results of TST-DL and PPB-DL.

The results show that the performance of both TST-DL and PPB-DL suffer from the increase of the compression ratio (model ill-posedness). This is observed as an increase of the RMSE and a decrease of the SSIM of the reconstructed images. This is reasonable since the model ill-posedness will affect both the initial guess images using the physics priors in PPB-DL and the learned inverse model in step 1 of TST-DL. However, TST-DL is more robust to the model ill-posedness since it starts to perform slightly better than PPB-DL at the 16X compression ratio (high model ill-posedness) compared with a slightly worse performance than PPB-DL at 2X and 4X compression ratios (low model ill-posedness). In TST-DL, there is more flexibility to incorporate features from the images in the training process. This makes the image reconstruction results more robust to highly ill-posed inverse models. However, in PPB-DL, the physics priors used for the initial LSQR image reconstruction are not able to incorporate sufficient information from the training dataset to offset the high compression of the image acquisition.

The results also show that both the performances of TST-DL and PPB-DL drop with the increase of the noise level (decrease of the SNR). However, they both have a similar good level of robustness to noise. For instance, at the 4X compression ratio, the results from TST-DL and PPB-DL still remain at a reasonable level with the RMSE lower than 0.11 and SSIM larger than 0.50 at the -5 dB SNR level (Figs. 2(e) and 2(f) show the 1D measurement data at the 4X compression ratio without noise and with -5 dB SNR of noise, respectively). Figures 2(h)–2(w) show a set of reconstructed images from TST-DL (FCL and final results) and PPB-DL (inputs and final results) with the 4X compression ratio at -5 dB, 0 dB, 10 dB SNR levels and the noise-free

Optics EXPRESS

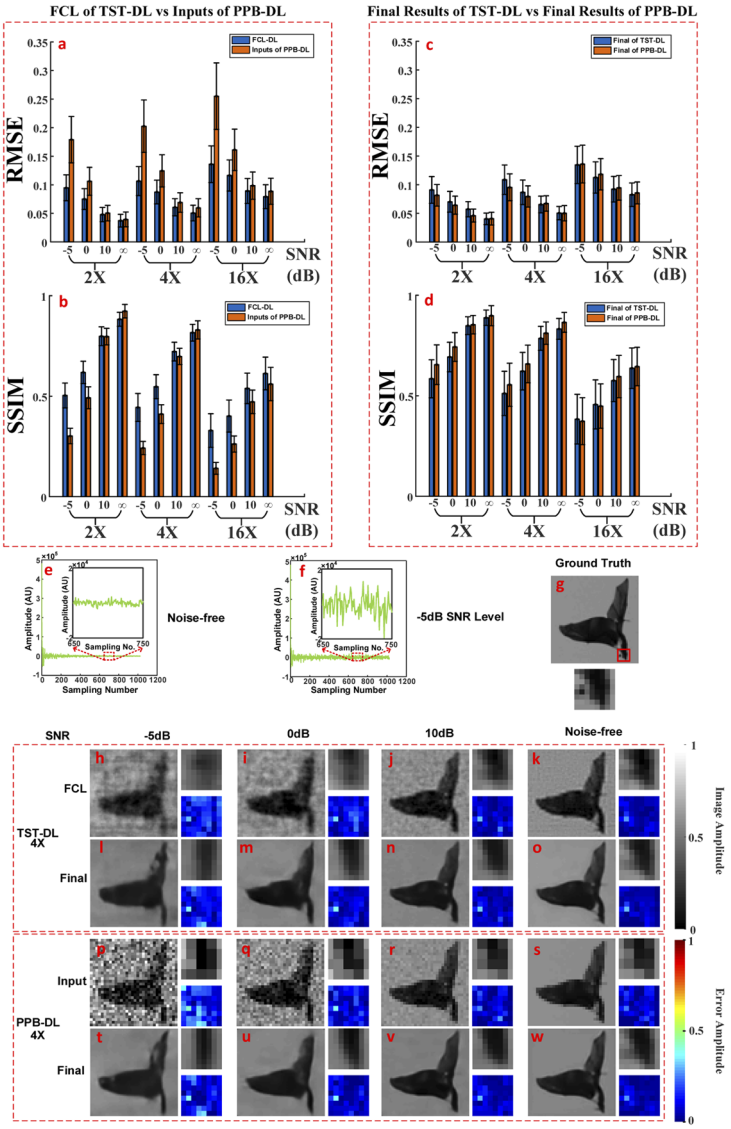

Fig. 2. TST-DL's Robustness to model ill-posedness and noise compared with PPB-DL at the 2X, 4X and 16X compression ratios with varying SNR levels of noise (-5 dB, 0 dB, 10 dB and the noise-free case). (a) RMSE and (b) SSIM of the intermediate reconstructed images from the first-step training in TST-DL (FCL of TST-DL) and initial image guesses as the inputs of PPB-DL (Inputs of PPB-DL). (c) RMSE and (d) SSIM of the final results in both TST-DL and PPB-DL. (e) 1D measurement data without noise at the 4X compression ratio. (f) 1D measurement data at the -5 dB SNR level at the 4X compression ratio. (g) The ground-truth of an image in the testing dataset and the fine detail in the red square. (h)-(k) FCL (of TST-DL) predictions of the image, the predictions of the fine detail in (g) and the prediction errors of the fine detail. (l)-(o) The final TST-DL predictions of the image, the predictions of the fine detail in (g) and the prediction errors of the fine detail. (p)-(s) The initial image guesses (as the inputs of PPB-DL), the fine detail in (g) and the errors of the fine detail. (t)-(w) The PPB-DL predictions of the image, the predictions of the fine detail in (g) and the prediction errors of the fine detail. The error bars represent the standard deviation of the RMSE or SSIM of the testing images with respect to the ground truth.

Optics EXPRESS

case with the same ground-truth image in Fig. 2(g). Although the reconstructed images become more and more blurred as the noise level increases, the general shape and even some of the details (the mouth of the bird in the zoom-in figures) can still be well reconstructed at the -5 dB, 0 dB and 10 dB SNR levels. Given the SNR levels at -5 dB and 0 dB are extremely high levels of noise (for 0 dB, the noise level is the same as the signal level), we can conclude that TST-DL is robust to noise.

The results in Fig. 2 show the inverse model learned in the FCL of TST-DL is robust to noise and ill-posedness. As seen in Figs. 2(a) and (b), the initial image estimate provided by the FCL outperforms the physics-based preprocessing in all but the most ideal case (2X compression ratio, noise-free). As the SNR decreases and the model ill-posedness increases, the advantage of the FCL becomes more evident. The FCL is capable to outperform LSQR (which incorporates perfect knowledge of the imaging model) because it incorporates data priors in the training process. In this particular case, the advantage disappears after the second step of training, but it does indicate that a robust inverse is learned in the FCL at all noise levels and degrees of ill-posedness tested here. Additional evidence that the FCL is capable and robust to estimate the inverse model are shown in detail in Supplement 1, Section 2.

Overall, these results show that TST-DL is comparable to PPB-DL which incorporates perfect knowledge of the physics priors. The physics priors are most beneficial when the compression ratio is small. As the compression ratio increases, it becomes advantageous to apply TST-DL, which can incorporate both the physical model and data priors into the first step of training.

3.2. Advantages of the two-step training strategy

As shown in Supplement 1, Section 1, the TST-DL approach outperforms OST-DL, in which an identical network is trained in a single step. We hypothesized that the improved performance comes from the added constraints applied by the two-step training strategy, which limits the effects of over-parameterization. In the TST-DL case, we constrained the FCL to be a good approximation of the inverse model by training it alone; no such constraint was placed on OST-DL. In order to investigate this, we analyzed the intermediate image that is produced between the FCL and the U-Net in Fig. 3. We applied both the two-step (TST-DL) and the one-step (OST-DL) training strategies to the reconstruction of simulated single-pixel camera images with the RD Hadamard matrix with varying compression ratios and SNR. In the TST-DL case, the FCL generates a good approximation of the image, as shown in Figs. 3(i) and 3(o) with the corresponding RMSE and SSIM quantified in Fig. 3(a) and (b). Then, the U-Net effectively denoises and regularizes the estimate, as shown in Figs. 3(j) and 3(p) with the corresponding RMSE and SSIM shown in Figs. 3(c) and 3(d). In the OST-DL case, the image after the FCL bears little resemblance to the ground truth as shown in Figs. 3(f) and 3(l) with a much higher RMSE and lower SSIM shown in Figs. 3(a) and 3(b). Thus, the FCL in OST-DL learns something other than the inverse of the physical imaging model. The U-Net part in OST-DL does a good job of completing the image reconstruction process as shown in Figs. 3(c) and 3(d) and in Figs. 3(g) and 3(m), but it is not quite able to achieve the performance of the TST-DL approach. The lack of additional constraints leads to a greater degree of overfitting that occurs in the OST-DL case. This is evident in the deviation between the training and validation losses during training (Figs. 3(h) and 3(n)). The overfitting becomes more pronounced at higher levels of noise.

Another possible contributing factor to the improved performance is the vanishing gradient problem. During backpropagation, the weights in earlier layers of a network have a smaller gradient than those in the later layers, leading to slow training or plateauing at suboptimal values. Alternative approaches, such as adding an auxiliary loss function after the FCL could help alleviate this problem [42] and a comparison with the TST-DL approach is made in details in Supplement 1, Section 5 with three imaging cases. The results in Supplement 1, Section 5 show

Optics EXPRESS

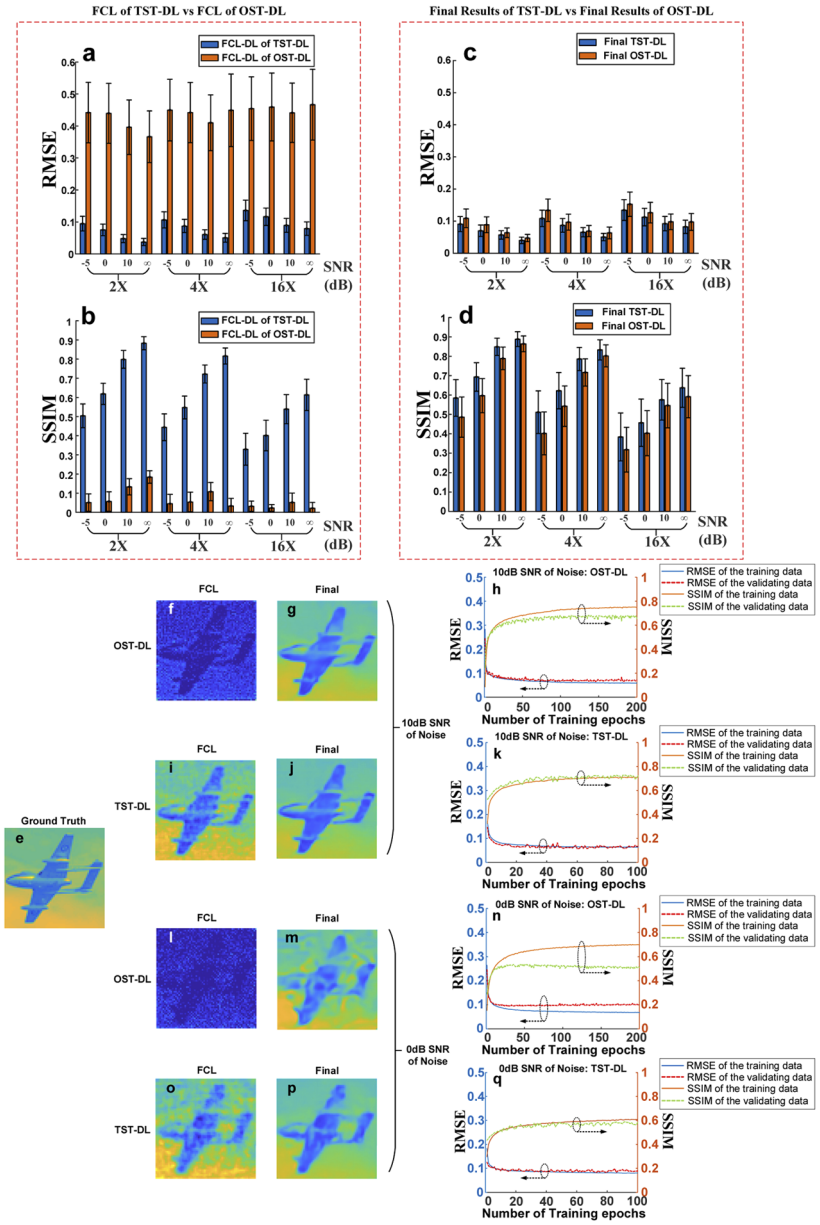

Fig. 3. Comparison between TST-DL and OST-DL with the RD Hadamard matrix at the 2X, 4X and 16X compression ratios with varying SNR levels of noise (-5 dB, 0 dB, 10 dB and the noise-free case). (a) RMSE and (b) SSIM of the intermediate results after the FCL in both TST-DL and OST-DL. (c) RMSE and (d) SSIM of the final results in both TST-DL and OST-DL. (e) The ground truth of a representative image. (f) The intermediate reconstructed image after the FCL, and (g) the final reconstructed image in OST-DL at the 4X compression ratio in the 10 dB SNR-of-noise case. (h) The RMSE and SSIM of the reconstructed images from both the training and validating data during the training process in OST-DL at the 4X compression ratio in the 10 dB SNR-of-noise case for overfitting analysis. (i)-(k) Are the same as (f)-(h) except for the TST-DL approach. (l)-(q) Are the same as (f)-(k) except for the 0-dB SNR case. The error bars represent the standard deviation of the RMSE or SSIM of the testing images with respect to the ground truth.

Optics EXPRESS

that, in general, TST-DL outperforms OST-DL with an auxiliary loss function (Multi-outputs OST-DL).

Overall, these findings indicate that the network architecture of OST-DL is complex enough to learn something other than the physical imaging model. In the absence of additional constraints, it tends to do just that.

3.3. Robustness to model mismatch

A preprocessing step can be helpful to the DL prediction by forming an initial estimate of the image as the input of the DL network [2]. Here, we sought to explore on under what circumstances a neural-network-based preprocessor should be used rather than the physics-prior-based preprocessor. We propose that the model mismatch (or uncertainty) and the noise are two key factors in the decision. The model mismatch is defined as the difference between the idealized model informed by the physics priors and the actual experimental model. We investigated the effects of the model mismatch in two commonly employed DL training and testing strategies in the image reconstruction of single-pixel imaging with the RD Hadamard patterns. The first strategy is to acquire both training and testing data from experiments. In this case, the model mismatch arises because, although the training data are generated with the actual experimental model, the idealized model is used to inform the physics prior in the preprocessing step. The second strategy is to acquire the training data from simulations and the testing data from experiments. In this case, the model mismatch is more detrimental because the training data and preprocessor both make use of the idealized model, which allows the errors to propagate all the way through the training process. In both cases, the actual experimental model is used to generate the testing data. The TST-DL was trained and tested using the first strategy since the model is unknown and directly learned. For the PPB-DL approach both strategies were used (termed PPB-DL and PPB-DL2, respectively).

Two specific types of the defined model mismatches were generated here. In the first case we randomly inverted a subset of the elements in the RD Hadamard matrix with a 4X compression ratio. For the second model mismatch we added different levels of uncertainty (as modeled by additive Gaussian random variables) to the RD Hadamard matrix with a 4X compression ratio. In both types of model mismatches, the modified matrix is equivalent to the actual experimental model and the original unmodified matrix is equivalent to the idealized model mentioned in the previous paragraph. For image reconstruction using PPB-DL and PPB-DL2, we still used the original unmodified matrix for an initial guess of the image as the input of PPB-DL and PPB-DL2. In order to explore the effects of noise while the model mismatch exists, noise levels of 15 dB and 0 dB SNR were used for both the training and testing measurement data. 15 dB SNR of noise is a reasonable noise level in an actual imaging system while 0 dB SNR of noise (the noise level equals to the mean signal level) was used as an extreme example. In PPB-DL2, the noise-free training measurement data were used for training and the noisy testing data were used for testing.

Figure 4 shows the results in TST-DL, PPB-DL, and PPB-DL2 with each type of the model mismatch. The results show that as the model mismatch increases, the performance of TST-DL, PPB-DL and PPB-DL2 all decrease. In the case of TST-DL, this decrease in performance stems from the fact that the actual forward model deviates from the more-ideal RD Hadamard matrix. However, the degree of decrease is different among TST-DL, PPB-DL and PPB-DL2. While the PPB-DL and PPB-DL2 approaches slightly outperform TST-DL when no model mismatch exists, their advantage begins to wane as the model mismatch grows. When a high degree of model mismatch occurs, then the TST-DL approach clearly outperforms the approaches which incorporate incorrect physics priors. Interestingly, this advantage becomes less pronounced as the noise level increases. Overall, the PPB-DL2 preforms much worse than PPB-DL and TST-DL because the training does not correct for the model mismatch introduced by the preprocessor.

Optics EXPRESS

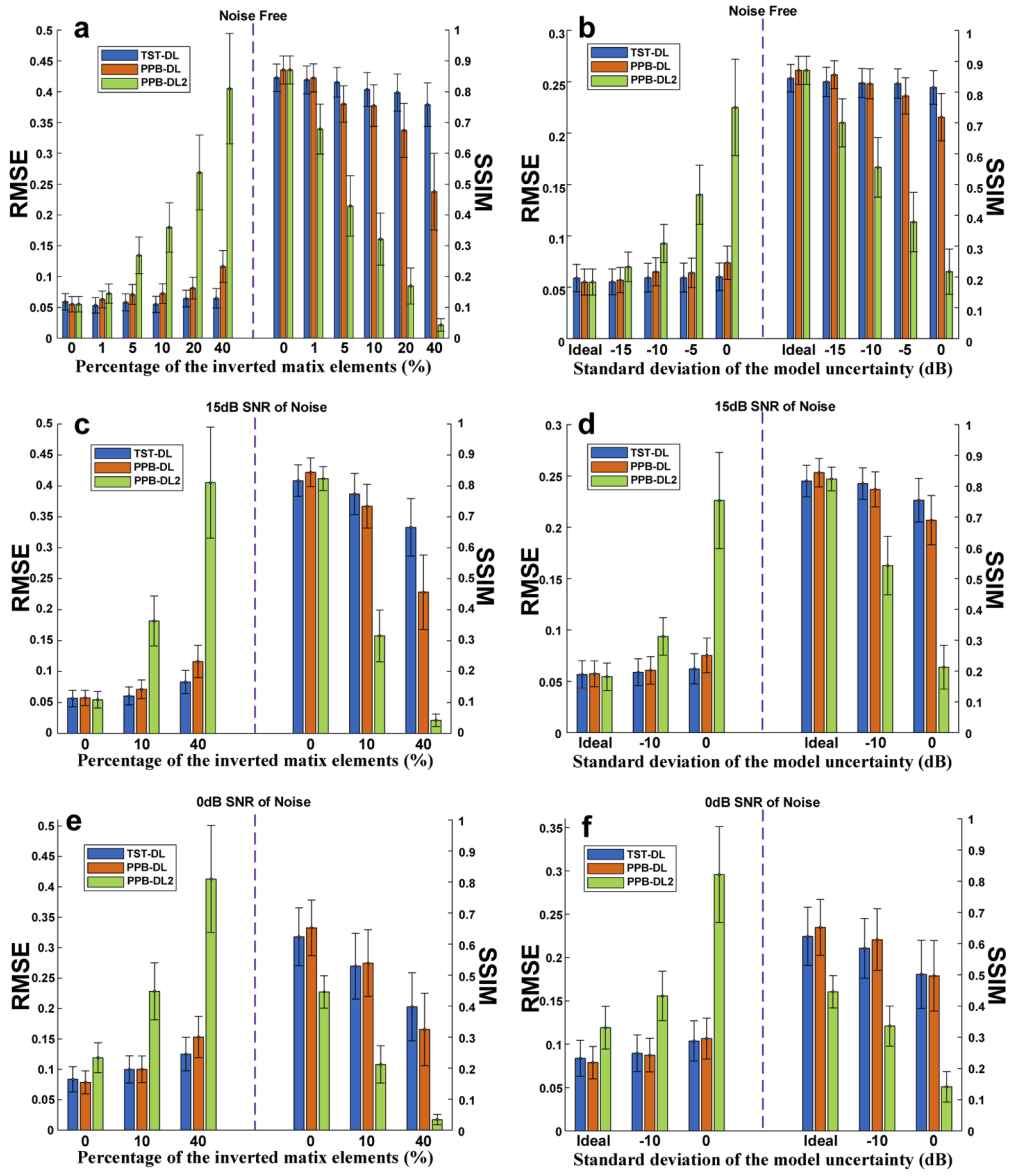

Fig. 4. Robustness of TST-DL, PPB-DL and PPB-DL2 to the model mismatch in the varying SNR cases. (a) RMSE and SSIM of the predicted results with varying percentages of the inverted matrix elements in the noise-free case. (b) RMSE and SSIM of the predicted results with Gaussian random variables with varying standard deviation added to the model in the noise-free case. (c) RMSE and SSIM of the predicted results with varying percentages of the inverted matrix elements in the 15 dB SNR of noise case. (d) RMSE and SSIM of the predicted results with Gaussian random variables with varying standard deviation added to the model in the 15 dB SNR of noise case. (e) RMSE and SSIM of the predicted results with varying percentages of the inverted matrix elements in the 0 dB SNR of noise case. (f) RMSE and SSIM of the predicted results with Gaussian random variables with varying standard deviation added to the model in the 0 dB SNR of noise case. The error bars represent the standard deviation of the RMSE or SSIM of the testing images with respect to the ground truth.

Optics EXPRESS

Taken as a whole, these results yield insight into when and how physics priors should be incorporated into a DL strategy. The TST-DL approach performed better when there is uncertainty in the model and there is a little-to-moderate amount of noise. As the noise approaches extreme levels (SNR = 0 dB), then it becomes beneficial to incorporate a preprocessor with relatively accurate physics priors. It is important to note that both cases (TST-DL and PPB-DL) rely on training with data produced without the model mismatch (i.e., the model mismatch is only present in the preprocessor step of PPB-DL). If the model is trained on data produced with the model mismatch (i.e., PPB-DL2), then even small levels of model uncertainty can lead to large errors. The TST-DL and PPB-DL approaches come with a drawback, however, that the training process must be performed with experimentally acquired images.

4. Experimental single-pixel imaging results

Experimentally recorded data in single-pixel imaging with random grayscale illumination patterns were used to verify the effectiveness of TST-DL (see Section 2.4.2 for model description and data acquisition). Figure 5 shows representative ground-truth images from the testing dataset as well as the corresponding reconstructed images from TwIST, PPB-DL, OST-DL and TST-DL at the 16X compression ratio. Qualitatively, both the PPB-DL and TST-DL approaches achieve better results than TwIST and OST-DL where the results from TwIST are prone to blurry and the results from OST-DL are prone to reconstrucing the wrong number. Quantitative comparison was made by calculating the mean and the standard deviation of RMSE and SSIM between the final reconstructed images and the ground-truth images in the testing dataset as shown in Fig. 5(k). The T-Test on the sets of RMSE and SSIM of TwIST, PPB-DL, OST-DL and TST-DL at the 16X compression ratio is shown in Supplement 1, Table S1. The quantitative comparison and T-Test show that TST-DL performs better than TwIST, OST-DL and PPB-DL at the 16X compression ratio, with a lower RMSE and higher SSIM. There are two possible reasons that TST-DL outperforms TwIST and PPB-DL. First, TST-DL is more robust to model ill-posedness than TwIST and PPB-DL since it has more flexibility to incorporate features from the training images. The initial LSQR image reconstruction in PPB-DL has poor image quality and insufficient information is available from the training dataset to offset the high compression of the image acquisition (as detailed in Section 3.1). Second, the mismatch between the idealized model (physics priors used in TwIST and PPB-DL) and the actual experimental model degrades the performances of TwIST and PPB-DL (as detailed in Section 3.3). A possible reason that TST-DL outperforms OST-DL is that the added constraints applied by TST-DL limit the effects of over-parameterization while OST-DL runs into the over-parameterization issue since no constraint is applied (as detailed in Section 3.2). The details and results for the 256X compression-ratio case are shown in Supplement 1, Section 4. The RMSE, SSIM and the T-Test results are included. According to the results, PPB-DL shows better contrast, but is also prone to reconstructing the wrong number. OST-DL is prone to both blurry and reconstructing the wrong number. In contrast, TST-DL produces blurred results for the less-successful cases. Indeed, it is hard to pick the better DL approach among PPB-DL, OST-DL and TST-DL in the 256X compression-ratio case. It is important to note that the 256x compression ratio is extremely high; each 32×32 image was reconstructed from only four measurements. This makes it an incredibly difficult problem to solve, even for the sparse images in the MNIST database shown here. Therefore, the advantage of TST-DL over OST-DL by adding the constraint may vanish given such an extremely high compression ratio of 256X. However, it still shows the good performance of TST-DL since the results of TST-DL are comparable to those of PPB-DL which incorporates the physics priors at the 256X compression ratio.

Optics EXPRESS

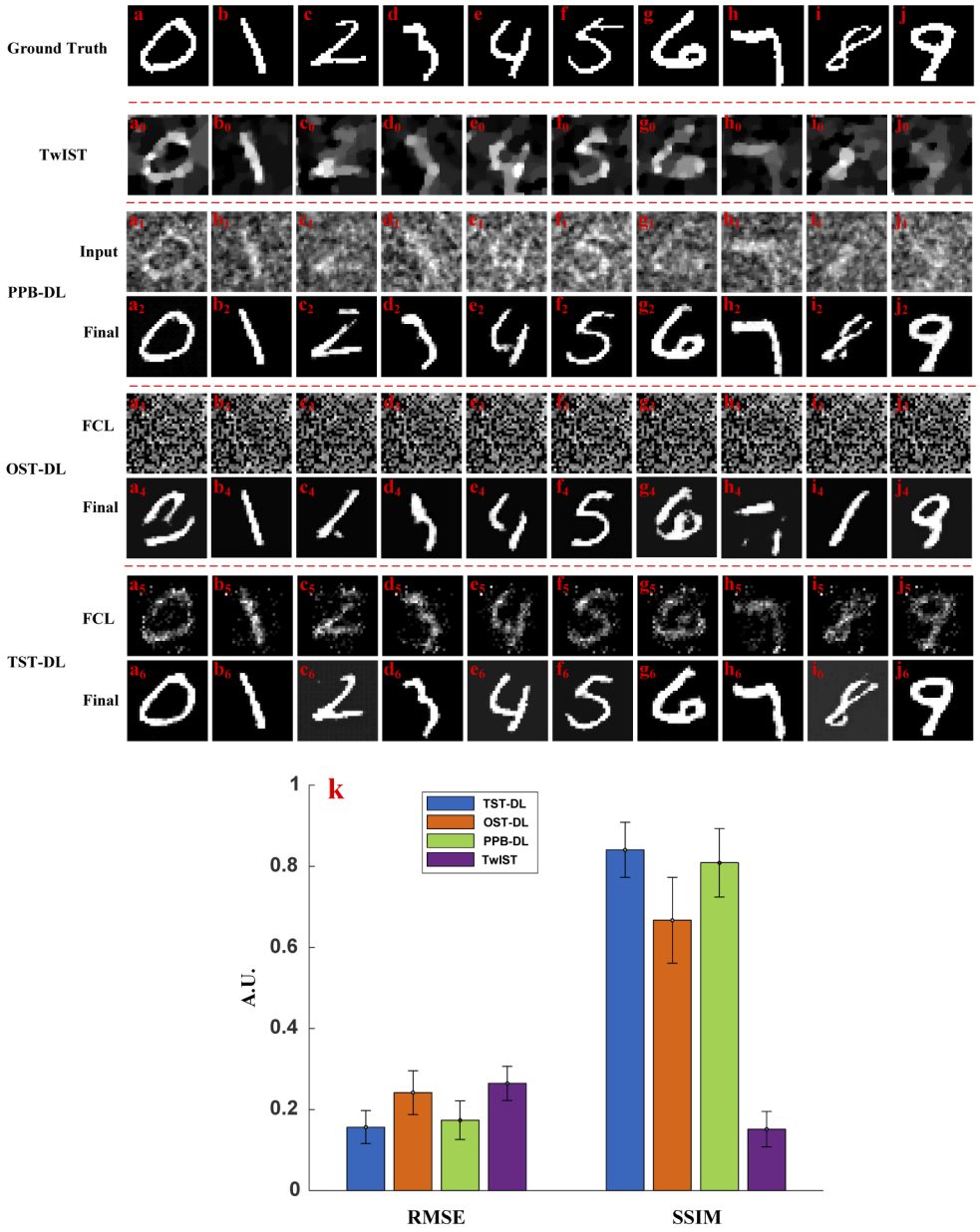

Fig. 5. Experimental results on single-pixel imaging with the 16X compression ratio. (a)-(j) Ground-truth images. (a_0)-(j_0) Reconstructed images in TwIST. (a_1)-(j_1) Initial image guesses as the inputs of PPB-DL. (a_2)-(j_2) Final reconstructed images in PPB-DL. (a_3)-(j_3) Intermediate images after the FCL in OST-DL. (a_4)-(j_4) Final reconstructed images in OST-DL. (a_5)-(j_5) Intermediate images after the FCL in TST-DL. (a_6)-(j_6) Final reconstructed images in TST-DL. (k) RMSE and SSIM between the final reconstructed images and the ground-truth images in the testing dataset for TST-DL, OST-DL, PPB-DL and TwIST. The error bars represent the standard deviation of the RMSE or SSIM of the testing images with respect to the ground truth.

Optics EXPRESS

5. Discussions and conclusions

A TST-DL framework is proposed for computational imaging without prior knowledge of the imaging model. The FCL in the first-step training acts as a preprocessor by directly learning the inverse of the forward operator given the training data. Then, the pre-trained FCL is fixed and concatenated with a U-Net for a second-step training as a regularization step. Simulations and experiments with different imaging models were conducted to verify the effectiveness of the proposed TST-DL with quantitative comparison with other DL frameworks and the iterative model-based optimization approaches. The results show the TST-DL outperforms the other DL frameworks without physics priors and is comparable to (and sometimes better than) the DL framework that incorporate the physics priors. The training time depends on the imaging model, the batch size and the number of training samples. For our simulated single-pixel imaging results with the 4X compression-ratio RD Hadamard patterns where the batch size is 50 with 10,000 training samples, each epoch in the first step of TST-DL took approximately 8 seconds and each epoch in the second step of TST-DL took approximately 17 seconds running on a NVIDIA Quadro M4000 GPU with an 8GB of memory. The average time to predict an image from the testing dataset in TST-DL is ≤ 1 ms in Tensorflow. A detailed comparison among the DL approaches in terms of the number of trainable parameters, the epoch number, flop counts and prediction time (ms per image) is shown in Supplement 1, Table S2.

The TST-DL framework is applicable to imaging problems for which a preprocessor is necessary. In applications where the acquired data and reconstructed images are more closely related, a standard U-Net (or other similar architecture) may be a more appropriate choice for end-to-end processing. Although this work focused primarily on a single-pixel imaging system, there are no model-specific restrictions to the approach. Thus, it could be readily adapted for a wide variety of imaging systems. It has the benefit of circumventing the model errors that arise from model-based preprocessors or simulated data. This comes with the drawback, however, that the training process must be performed with experimentally acquired images. This could be challenging since a relatively large dataset may be required to train the large number of parameters in the FCL as the size of the image increases, though the results are relatively robust as the dataset size is decreased (Supplement 1, Fig. S9).

The capability of TST-DL to handle nonlinear imaging models is analyzed in Supplement 1, Section 3 with the image de-autocorrelation problem as a test case. Further exploration is still needed in determining the optimal number of FCLs to use, the choice of the nonlinear activation functions in each FCL and the comparison with the existing DL approaches to solve such inverse problems [43–46]. This optimization will likely depend on the degree of nonlinearity of the model.

In summary, the TST-DL approach enables reliable image reconstruction without relying on possibly flawed assumptions about the imaging model. The two-step training strategy constrains the training process so that the inverse model can be effectively learned. Overall, this provides a flexible, standardized framework that can be applied to diverse imaging problems.

Funding. National Institutes of Health (R21GM137334); Neukom Institute for Computational Sciences (CompX Faculty Grant).

Acknowledgments. We thank Dr. Shuming Jiao and Miss Jun Feng (Nanophotonics Research Center, Shenzhen University, China) for the discussion on the single-pixel imaging experiment.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. J. N. Mait, G. W. Euliss, and R. A. Athale, "Computational imaging," Adv. Opt. Photonics **10**(2), 409–483 (2018).

Optics EXPRESS

2. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**(8), 921–943 (2019).
3. E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Probl. **23**(3), 969–985 (2007).
4. E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," IEEE Signal Process. Mag. **25**(2), 21–30 (2008).
5. L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**(7529), 74–77 (2014).
6. Z. Wang, L. Spinoulas, K. He, L. Tian, O. Cossairt, A. K. Katsaggelos, and H. Chen, "Compressive holographic video," Opt. Express **25**(1), 250–262 (2017).
7. R. Shang, R. Archibald, A. Gelb, and G. P. Luke, "Sparsity-based photoacoustic image reconstruction with a linear array transducer and direct measurement of the forward model," J. Biomed. Opt. **24**(03), 031015 (2018).
8. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**(7553), 436–444 (2015).
9. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**(11), 1437–1443 (2017).
10. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Sci. Appl. **7**(2), 17141 (2018).
11. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**(10), 1181–1190 (2018).
12. Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," Optica **6**(5), 618–629 (2019).
13. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**(9), 1117–1125 (2017).
14. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," Phys. Rev. Lett. **121**(24), 243902 (2018).
15. A. Goy, G. Rughoobur, S. Li, K. Arthur, A. I. Akinwande, and G. Barbastathis, "High-resolution limited-angle phase tomography of dense layered objects using deep neural networks," Proc. Natl. Acad. Sci. **116**(40), 19848–19856 (2019).
16. J. Liu, Q. He, and J. Luo, "A compressed sensing strategy for synthetic transmit aperture ultrasound imaging," IEEE Trans. Med. Imaging **36**(4), 878–891 (2017).
17. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241.
18. W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, E. X. Wu, and X. Tang, "Prostate segmentation using 2D bridged U-net," in *2019 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2019), pp. 1–7.
19. S. Shah, P. Ghosh, L. S. Davis, and T. Goldstein, "Stacked U-Nets: a no-frills approach to natural image segmentation," arXiv preprint arXiv:1804.10343 (2018).
20. J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *Advances in neural information processing systems* (2016), pp. 3036–3044.
21. T. Sun, Z. Chen, W. Yang, and Y. Wang, "Stacked U-Nets With Multi-Output for Road Extraction," in *CVPR Workshops* (2018), pp. 202–206.
22. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," Inverse Probl. Sci. Eng. **27**(7), 987–1005 (2019).
23. D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, and L. Maier-Hein, "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," in *Photons Plus Ultrasound: Imaging and Sensing 2018* (International Society for Optics and Photonics, 2018), p. 104942S.
24. G. Zhang, T. Guan, Z. Shen, X. Wang, T. Hu, D. Wang, Y. He, and N. Xie, "Fast phase retrieval in off-axis digital holographic microscopy through deep learning," Opt. Express **26**(15), 19388–19405 (2018).
25. T. Shimobaba, Y. Endo, T. Nishitsuji, T. Takahashi, Y. Nagahama, S. Hasegawa, M. Sano, R. Hirayama, T. Kakue, and A. Shiraki, "Computational ghost imaging using deep learning," Opt. Commun. **413**, 147–151 (2018).
26. D. Lee, J. Yoo, S. Tak, and J. C. Ye, "Deep residual learning for accelerated MRI using magnitude and phase networks," IEEE Trans. Biomed. Eng. **65**(9), 1985–1995 (2018).
27. H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu, "Image prediction for limited-angle tomography via deep learning with convolutional neural network," arXiv preprint arXiv:1607.08707 (2016).
28. J. Schwab, S. Antholzer, R. Nuster, G. Paltauf, and M. Haltmeier, "Deep Learning of truncated singular values for limited view photoacoustic tomography," in Photons Plus Ultrasound: Imaging and Sensing 2019 (International Society for Optics and Photonics, 2019), p. 1087836.
29. S. Guan, A. Khan, S. Sikdar, and P. Chitnis, "Fully Dense UNet for 2D sparse photoacoustic tomography artifact removal," *IEEE journal of biomedical and health informatics* (2019).
30. C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," ACM Trans. Math. Softw. **8**(1), 43–71 (1982).
31. 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**(12), 2992–3004 (2007).

Optics EXPRESS

32. B. I. Erkmen, "Computational ghost imaging for remote sensing applications," IPN Progress Report 42, 1–23 (2011).
33. F. Wang, H. Wang, H. Wang, G. Li, and G. Situ, "Learning from simulation: An end-to-end deep-learning approach for computational ghost imaging," Opt. Express **27**(18), 25560–25572 (2019).
34. M. Deng, S. Li, Z. Zhang, I. Kang, N. X. Fang, and G. Barbastathis, "On the interplay between physical and content priors in deep learning for computational imaging," Opt. Express **28**(16), 24152–24170 (2020).
35. C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, "Deep learning for real-time single-pixel video," Sci. Rep. **8**(1), 2369 (2018).
36. S. Jiao, J. Feng, Y. Gao, T. Lei, Z. Xie, and X. Yuan, "Optical machine learning with incoherent light and a single-pixel detector," Opt. Lett. **44**(21), 5186–5189 (2019).
37. S. Jiao, Y. Gao, J. Feng, T. Lei, and X. Yuan, "Does deep learning always outperform simple linear regression in optical imaging?" Opt. Express **28**(3), 3717–3731 (2020).
38. M.-J. Sun, L.-T. Meng, M. P. Edgar, M. J. Padgett, and N. Radwell, "A Russian Dolls ordering of the Hadamard basis for compressive single-pixel imaging," Sci. Rep. **7**(1), 3464 (2017).
39. A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (2011), pp. 215–223.
40. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**(11), 2278–2324 (1998).
41. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**(4), 600–612 (2004).
42. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*(2015), pp. 1–9.
43. C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," Optica **7**(1), 63–71 (2020).
44. M. Liao, S. Zheng, D. Lu, G. Situ, and X. Peng, "Real-time imaging through moving scattering layers via a two-step deep learning strategy," in Unconventional Optical Imaging II (International Society for Optics and Photonics, 2020), p. 113510V.
45. E. Guo, S. Zhu, Y. Sun, L. Bai, C. Zuo, and J. Han, "Learning-based method to reconstruct complex targets through scattering medium beyond the memory effect," Opt. Express **28**(2), 2433–2446 (2020).
46. M. Lyu, H. Wang, G. Li, S. Zheng, and G. Situ, "Learning-based lensless imaging through optically thick scattering media," Adv. Photonics **1**(3), 036002 (2019).