Unrolled, model-based networks for lensless imaging

Anonymous Author(s)
Affiliation
Address
email

Abstract

We develop end-to-end learned reconstructions for lensless mask-based cameras, including an experimental system for capturing aligned lensless and lensed images as a test platform. Various reconstruction methods are explored, on a scale from classic iterative approaches (based on the physical imaging model) to deep learned methods with many learned parameters. In the middle ground, we present several variations of unrolled alternating direction method of multipliers (ADMM), each with varying numbers of learned parameters. The network structure combines knowledge of the physical imaging model with learned parameters updated from the data, which compensate for artifacts caused by physical approximations. This unrolled approach is \(20\times\) faster than classic methods and produces better visual reconstruction quality than both the classic and deep methods on our experimental system.

1 Introduction

Mask-based lensless cameras can be small, light-weight, and capture higher-dimensional information, such as 3D and video, from a single shot \([1,5,8,13]\). Instead of using a lens, lensless cameras use a phase or amplitude mask which maps points in the world to a unique multiplexed pattern on the sensor (Fig. 1(a)). Typically, a reconstruction method based on convex optimization is used to iteratively solve for the scene from the multiplexed sensor data. In practice, iterative methods can be slow and the reconstruction quality is sensitive to errors from model mismatch, imperfect calibration, hand-tuned parameters, and hand-picked priors which are not necessarily representative of the data. Solving the inverse problem with deep methods offers a favorable alternative due to the decreased computation and the ability to directly optimize image quality. However, this comes at the price of thousands of training pairs, a loss of interpretability, and the inability to explicitly add prior knowledge, such as the imaging system physics, into the network.

Unrolled optimization has emerged as a promising middle-ground approach between classic and deep methods for a variety of inverse problems \([5,7,11,12]\). In unrolled optimization, a fixed number of iterations from a classic algorithm is interpreted as a deep network, with each iteration serving as a layer in the network. In each layer, if the parameters of the algorithm are differentiable with respect to the output, they can be optimized for a given loss function through backpropagation. Following this framework, we unroll the iterative alternating direction method of multipliers (ADMM) algorithm with a variable splitting specific for lensless imaging \([1,4]\). This allows us to incorporate knowledge of the image formation process into the neural network as well as directly optimize image reconstruction quality based on training examples. We present several variations of networks along the spectrum between classic and deep methods, by varying the number of trainable parameters and test the networks on a real imaging system.

To train our networks, we experimentally capture a large dataset of 25,000 aligned lensed and lensless images as shown in Fig. 1(a). Our full dataset and models are publicly available. We demonstrate a \(20\times\) speedup and \(3\times\) improvement in perceptual similarity for lensless imaging reconstructions, showing that our unrolled method outperforms both the classic and deep approaches in terms of

2 Methods

To formulate our unrolled network, we first describe our imaging forward model. Our lensless imaging model can be approximated as a cropped convolution between the scene and the point spread function (PSF) of the system. In practice, the PSF is measured experimentally using an LED placed at the desired focal distance of the system. Assuming all points in the scene are incoherent with each other, our sensor measurement, \( b \), can be described as:

\[
b(x, y) = \text{crop}[h(x, y) \ast x(x, y)] = CHx,
\]

where \( h \) is the system PSF, \( x \) represents the scene, and \((x, y)\) are the sensor coordinates. Here, \( \ast \) denotes 2D discrete linear convolution, which returns an array that is larger than both the scene and the PSF. Therefore, a crop operation restricts the output to the physical sensor size. This relation is represented compactly in matrix-vector notation with crop denoted as \( C \) and convolution with the PSF denoted as \( H \).

To efficiently solve the inverse problem, we use ADMM with a variable splitting that leverages the structure of the problem. The inverse problem is formulated as:

\[
\hat{x} = \arg\min_{w \geq 0, u, v} \frac{1}{2}\|b - Cw\|_2^2 + \tau\|u\|_1, \\
s.t. v = Hx, u = \Psi x, w = x,
\]

where \( \Psi \) is a sparsifying transform, such as finite differences for total variation (TV) denoising, and \( \tau \) is a tuning parameter that adjusts the sparsity level. The update equations for each iteration become:

Figure 1: (a) Experimental Setup. We display images on a computer screen and use a beamsplitter to simultaneously record measurements on both a lensed and lensless camera. (b) Unrolled network architecture. The input measurement and the calibration PSF are first fed into \( N \) layers of unrolled ADMM. At each layer, the updates corresponding to Eq. (3) are applied. The output of this can be fed into an optional denoiser network such as a U-Net \[10\]. The network parameters are updated based on a loss function comparing the output image to the lensed image. Red arrows represent backpropagation through the network parameters.
We analyze three variations of unrolled ADMM, each having a different number of learned parameters, without the beamsplitter and computer monitor. Le-ADMM also produces visually appealing images, addressing two of the big limitations of lensless imagers. Our learned network is fast enough for interactive previewing of the scene and has the most learned parameters out of our networks and also has the best performance, achieving a comparably to converged ADMM, with the best networks operating at 20\times the speed. Le-ADMM-U has the most learned parameters out of our networks and also has the best performance, achieving a better MSE and LPIPS score than traditional ADMM as well as the U-Net. Figure 3 shows several sample reconstructions from our test set as well as some reconstructions of images taken in the wild without the beamsplitter and computer monitor.

Our work presents a preliminary analysis of using unrolled, model-based neural networks on a real experimental lensless imaging system. We show that it is favorable to use a network that incorporates knowledge of the imaging system physics, along with trainable parameters to optimize the network performance. We can perform comparably to classic algorithms at a fraction of the speed using only a few learned parameters, but can greatly improve image quality when increasing the number of learned parameters. Our learned network is fast enough for interactive previewing of the scene and also produces visually appealing images, addressing two of the big limitations of lensless imagers.

\begin{equation}
\begin{aligned}
\mathbf{x}^{k+1} &\leftarrow (\mu_3^k \mathbf{H}^\top \mathbf{H} + \mu_2^k \Psi^\top \Psi + \mu_1^k I)^{-1} \mathbf{v}^k \\
\mathbf{v}^{k+1} &\leftarrow (\mathbf{C}^\top \mathbf{C} + \mu_1^k I)^{-1} (\mathbf{C}^\top \mathbf{H} \mathbf{x}^k + \mathbf{C}^\top \mathbf{b}) \\
\mathbf{w}^{k+1} &\leftarrow \max(\mathbf{Hx}^k + \mathbf{z}^k, 0) \\
\mathbf{u}^{k+1} &\leftarrow \mathcal{T}_\alpha(\Psi(\mathbf{x}^k) + \alpha_2^k / \mu_2^k) \\
\alpha_1^{k+1} &\leftarrow \alpha_1^k + \mu_2^k (\mathbf{Hx}^{k+1} - \mathbf{v}^{k+1}) \\
\alpha_2^{k+1} &\leftarrow \alpha_2^k + \mu_2^k (\Psi(\mathbf{x}^{k+1}) - \mathbf{u}^{k+1}) \\
\alpha_3^{k+1} &\leftarrow \alpha_3^k + \mu_3^k (\mathbf{x}^{k+1} - \mathbf{w}^{k+1})
\end{aligned}
\end{equation}

Here, \(\alpha_1, \alpha_2, \) and \(\alpha_3\) are the Lagrange multipliers, or dual variables, respectively associated with \(u, v, w, \) and \(\mu_1, \mu_2, \) and \(\mu_3\) are scalar penalty parameters. \(\mathcal{T}_\alpha\) denotes vectorial soft-thresholding with parameter \(\tau / \mu_2.\) To unroll the network, we model each \(k\)th iteration of ADMM as a layer in a neural network. We denote the collection of update equations at the \(k\)th step of ADMM as \(S^k.\)

We analyze three variations of unrolled ADMM, each having a different number of learned parameters, denoted by \(\Theta.\) These networks are visualised in Figure 3(b). The three variations are summarized as:

- **Le-ADMM** (20 parameters, \(\Theta = \{\mu_1^k, \mu_2^k, \mu_3^k, \tau^k\}\) - Learned ADMM has trainable tuning and hyper-parameters.
- **Le-ADMM*** (32,135 parameters, \(\Theta = \{\mu_1^k, \mu_2^k, \mu_3^k, \mathcal{N}\}\) - extends Le-ADMM by adding a trainable convolutional neural network (CNN) instead of a hand-tuned sparsifying transform. \(\mathcal{N}\) represents a learnable network and replaces the \(\mathbf{v}^{k+1}\) update of Eq. (3).
- **Le-ADMM-U** (10,605,927 parameters, \(\Theta = \{\mu_1^k, \mu_2^k, \mu_3^k, \tau^k, \mathcal{U}\}\) - adds a trainable deep denoiser based on a CNN as the last layer of the Le-ADMM network, learning both the hyper-parameters of Le-ADMM as well as the denoiser.

For training, we simultaneously collect a set of lensless and ground truth image pairs using an experimental setup with a lensed camera, lensless camera, beamsplitter, and computer monitor as shown in Fig. 1(a). We use two Basler Dart (daA1920-30uc) sensors, one with a 6mm focal length S-mount lens (lensend), and one with off-the-shelf phase mask (Luminit 0.5\%\) and laser-cut aperture (lensless). To achieve pixel-wise alignment between the image pairs, we first optically align the two cameras, then perform a digital calibration process to co-align both cameras’ coordinate systems. We capture 25,000 images from the MirFlickr dataset [9]. After down-sampling and cropping, the final dataset images are 380\times 210 pixels. We separate these into 24,000 training images and 1,000 test images. We use a combination of mean-squared error (MSE) and LPIPS from [14] for training.

3 Results

Figure 2 summarizes the performance of our networks on the test set. We compare against ADMM run to convergence (100 iterations), ADMM bounded to 5 iterations (similar run time to our unrolled network), as well as against an end-to-end trained U-Net [10]. All of our networks perform better or comparably to converged ADMM, with the best networks operating at 20\times the speed. Le-ADMM-U has the most learned parameters out of our networks and also has the best performance, achieving a better MSE and LPIPS score than traditional ADMM as well as the U-Net. Figure 3 shows several sample reconstructions from our test set as well as some reconstructions of images taken in the wild without the beamsplitter and computer monitor.

Our work presents a preliminary analysis of using unrolled, model-based neural networks on a real experimental lensless imaging system. We show that it is favorable to use a network that incorporates knowledge of the imaging system physics, along with trainable parameters to optimize the network performance. We can perform comparably to classic algorithms at a fraction of the speed using only a few learned parameters, but can greatly improve image quality when increasing the number of learned parameters. Our learned network is fast enough for interactive previewing of the scene and also produces visually appealing images, addressing two of the big limitations of lensless imagers.
Figure 2: Network Performance on test set. On average, reconstructions from our learned networks (green) are more similar to the ground truth lensed images (lower MSE and LPIPS) than those from 5 iterations of ADMM. Furthermore, our networks have comparable or better performance than bounded ADMM (100 iterations), which takes $20 \times$ longer than Le-ADMM and Le-ADMM-U. The data fidelity term is higher for the learned methods, indicating that these reconstructions are less consistent with the image formation model. This suggests that the models are compensating for our simple forward model and are able to produce higher quality images.

Figure 3: Reconstruction results for methods along the spectrum from classic to deep, with the raw lensless measurement (contrast stretched) and the ground truth images from the lensed camera for reference. Le-ADMM has similar image quality to converged ADMM and better image quality than bounded ADMM (5 iter). Le-ADMM* and Le-ADMM-U have noticeably better visual image quality. The fully deep U-Net by itself is unable to reconstruct the appropriate colors and lacks detail.
References


