PhaseNet: A Deep Learning Based Phase Reconstruction Method for Ground-based Astronomy*

Dihan Zheng♮†, Shiqi Tang♮‡, Roland Wagner§, Ronny Ramlau¶, Chenglong Bao‖, and Raymond Chan#

Abstract. Ground-based astronomy utilizes modern telescopes to obtain information on the universe by analyzing recorded signals. Due to atmospheric turbulence, the reconstruction process requires solving a deconvolution problem with an unknown point spread function (PSF). The crucial step in PSF estimation is to obtain a high-resolution phase from low-resolution phase gradients, which is a challenging problem. In this paper, when multiple frames of low-resolution phase gradients are available, we introduce PhaseNet, a deep learning approach based on the Taylor frozen flow hypothesis. Our approach incorporates a data-driven residual regularization term, of which the gradient is parameterized by a network, into the Laplacian regularization based model. To solve the model, we unroll the Nesterov accelerated gradient algorithm so that the network can be efficiently and effectively trained. Finally, we evaluate the performance of PhaseNet under various atmospheric conditions and demonstrate its superiority over TV and Laplacian regularization based methods.

Key words. image deconvolution, astronomical imaging, machine learning, deep unrolling method.

MSC codes. 85-08, 68U10, 68T07

1. Introduction.

Observation of the universe is crucial in advancing scientific discoveries, and modern telescopes play a significant role in this perspective. In ground-based astronomy, images of objects in outer space are acquired via ground-based telescopes. However, the imaging system is generally affected by atmospheric turbulence, and the resulting images are usually blurred. As for any optical system, the observed image $g(\mathbf{x})$ of a ground-based astronomical telescope can be described as a convolution of the geometrical image $f(\mathbf{x})$ with

---

*Submitted to the editors DATE.

Funding: This work was funded by the Fog Research Institute under contract no. FRI-454. The work of DZ and CB was funded by National Key R&D Program of China under Grant 2021YFA1001300 and National Natural Science Foundation of China under Grant 12271291. The work of RR and RW was funded by the Austrian Science Fund (FWF), project F6805-N36, SFB Tomography Across the Scales. The work of ST and RC was funded by HKRGC GRF grants CityU1101120, CityU11309922 and CRF grant C1013-21GF.

†Yau Mathematical Sciences Center, Tsinghua University, Beijing, China (zhengdh19@mails.tsinghua.edu.cn).

‡Department of Mathematics, City University of Hong Kong, Kawloon Tong, Hong Kong (sqtang2-c@my.cityu.edu.hk).

§Industrial Mathematics Institute, Johannes Kepler University, Altenberger Strasse 69, 4040 Linz, Austria (wagner@indmath.uni-linz.ac.at).

¶Industrial Mathematics Institute, Johannes Kepler University, Altenberger Strasse 69, 4040 Linz, Austria and Johann Radon Institute for Computational and Applied Mathematics, Altenberger Strasse 69, 4040 Linz, Austria (ronny.ramlau@jku.at).

‖Yau Mathematical Sciences Center, Tsinghua University, Beijing, China, and Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing, China (clbao@mail.tsinghua.edu.cn)

#Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong. Hong Kong Center for Cerebro-Cardiovascular Health Engineering, Hong Kong Science Park, Hong Kong. (raymond.chan@cityu.edu.hk)

♮These authors contributed equally.

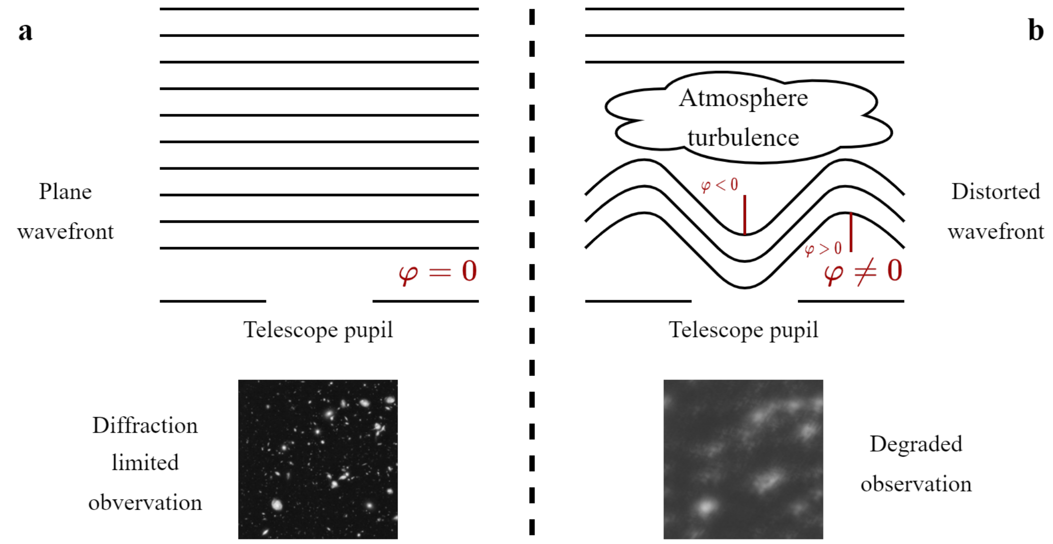


Figure 1: Illustrate the relation between wavefront and phase φ. a: no atmosphere distortion, corresponding to plane wavefront and $\varphi = 0$. b: wavefront is distorted by atmosphere turbulence, φ measures the intensity of the distortion.

the so-called Point Spread Function (PSF) $k(\mathbf{x}, \mathbf{y})$, *i.e.*,

$$g(\mathbf{x}, \mathbf{y}) = (k * f)(\mathbf{x}, \mathbf{y}) + n(\mathbf{x}, \mathbf{y}), \tag{1.1}$$

where $*$ denotes the convolution operator, $n(\mathbf{x}, \mathbf{y})$ models the noise. Since k is usually unknown as it is generated by the atmospheric turbulence and changes rapidly, one idea to solve (1.1) is to use image blind-deconvolution methods [7, 26, 21, 3]. However, the prior knowledge used in those methods, such as sparsity and smoothness, rarely holds for the blurring caused by the turbulent atmosphere [5]. In ground-based astronomy, the PSF arises from deviations in the incoming wavefront incident on the telescope. Ideally, without additional aberrations from the atmosphere or imperfections in the instrument, the wavefront is planar or has zero deviation. The resulting observation $g(\mathbf{x}, \mathbf{y})$ is called the diffraction-limited image, and the PSF of the telescope is given by

$$k(\mathbf{x}, \mathbf{y}) = |\mathcal{F}^{-1}(\mathcal{P})(\mathbf{x}, \mathbf{y})|^2, \tag{1.2}$$

which depends only on the pupil shape of the telescope, where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform and $\mathcal{P}$ is the aperture function of the telescope (1 inside the telescope aperture, 0 otherwise), see Figure 1a. However, in practical imaging, atmospheric turbulence degrades telescope image quality, leading to non-zero wavefront deviations quantified by the phase φ, see Figure 1b. Specifically, the phase φ of a wave measures its position in the cycle, indicating the oscillation level at a point. However, various optical imperfections can distort the wavefront from being ideally flat or spherical. These deviations are described by the phase φ, representing how far each point is ahead or behind the ideal wavefront. This phase function describes the actual wavefront shape. The phase measures delay: a positive value means the wavefront lags the reference, while a negative value means it leads. Thus, the phase map φ describes the wavefront deviation from the ideal shape at each point.

Using the Fourier optics model [2, 16], the PSF $k(\mathbf{x}, \mathbf{y})$ for an observation through atmospheric turbulence can be modeled as

$$k(\mathbf{x}, \mathbf{y}) = |\mathcal{F}^{-1}(\mathcal{P} \exp[\iota\varphi])(\mathbf{x}, \mathbf{y})|^2, \tag{1.3}$$

where $\iota = \sqrt{-1}$, and φ is the phase. The phase φ varies over time due to changes in atmospheric turbulence, with the short timescale over which φ is constant called the atmospheric coherence time τ_0. Therefore, the PSF in (1.3) is an instantaneous PSF. Adaptive Optics (AO) systems are developed to compensate for the effects of atmospheric turbulence, but residual aberrations remain. AO systems use wavefront sensors (WFS) to measure the incoming phase indirectly and one or more deformable mirrors (DM) to compensate for the observed distortions. Thus, (1.3) remains valid, but φ is the residual phase after AO compensation.

In most astronomical applications, the exposure time of the science camera is much longer than atmospheric coherence time τ_0. Consequently, the observed image is degraded by a time average over the resulting instantaneous PSFs [8]. Therefore, methods were developed to reconstruct this time-averaged PSF from saved AO telemetry data (see, e.g., [46] for an overview). Such a reconstructed PSF can then be used to improve the observed images, e.g., using recently developed methods tailored to ground-based astronomy developed in [12, 36].

WFSs split the incoming light into sub-apertures using a lenslet array, as shown in Figure 2. The intensity of light reaching the aperture is measured in photons. The number of incoming photons is limited, and each sub-aperture intercepts a small portion. If a WFS is designed as a fine grid, each sub-aperture may receive too few photons to provide a measurement that is not dominated by noise. Due to these physical limitations, the incoming phase can only be measured on a coarse grid, leading to a coarse approximation for the PSF. Therefore, recovering a high-resolution phase φ from the low-resolution WFS data can improve the reconstructed PSF and image restoration, which is the main goal of the work.

1.1. Problem modelling.

In Kolmogorov's theory [24, 47], the atmospheric turbulence φ is assumed to be a homogeneous and isotropic Gaussian process, which is assumed to be zero-centered with a covariance operator C_φ of the form:

$$C_\varphi = \mathcal{F}^{-1}\mathcal{M}\mathcal{F}, \tag{1.4}$$

where $\mathcal{M}$ is defined as $\mathcal{M}(f)(\kappa) = m(\kappa)f(\kappa)$ and m is known as the power spectrum. According to the Kolmogorov–Obukhov law of turbulence, the power spectrum m follows the power law:

$$m(\kappa) = C|\kappa|^{-11/3}, \quad L_{\text{in}} \le |\kappa| \le L_{\text{out}} \tag{1.5}$$

where $[L_{\text{in}}, L_{\text{out}}]$ is the inertial range, and C is a constant that measures the intensity of the turbulence. The singularity at zero in (1.5) makes expanding the power law outside the inertial range difficult. Therefore, we adopt the commonly used von Karman power spectral

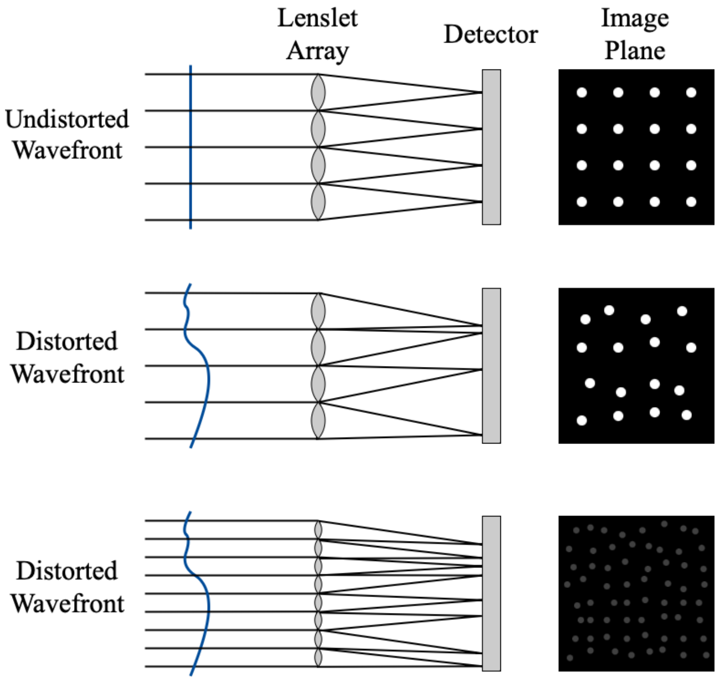


Figure 2: Principle of the Shack-Hartmann wavefront sensor: (top) undistorted wavefront, (middle) wavefront distorted by atmospheric turbulence, (bottom) distorted wavefront with fine-grid lenslet array.

density model [38]:

$$m(\kappa) = \frac{0.023 r_0^{-5/3}}{(\kappa_0^2 + |\kappa|^2)^{11/6}}, \tag{1.6}$$

where r_0 is the Fried parameter, $\kappa_0 = 1/L_{\text{out}}$, and L_{out} is the atmospheric turbulence outer-scale. The von Karman power law removes the singularity at zero and coincides asymptotically with (1.5) in the high-frequency region.

In practice, the atmosphere is composed of several layers that are located at different altitudes [37]. A geometric model describes wavefront propagation through turbulence, where the incoming wavefront φ is the sum of the wavefront passing through all turbulence layers. Assuming L turbulence layers, the wavefront φ decomposes orthogonally to the telescope direction as:

$$\varphi(\mathbf{x}) = \sum_{l=1}^{L} \varphi_l(\mathbf{x}), \tag{1.7}$$

where φ_l is the corresponding atmospheric turbulence in the l-th layer, and $\varphi_1, \ldots, \varphi_L$ are independent Gaussian processes with covariance operator $C_{\varphi_1}, \ldots, C_{\varphi_L}$.

The aberration φ can be measured by a Shack-Hartmann wavefront sensor (SH-WFS) [34]. Assuming the SH-WFS is composed of $m \times m$ sub-apertures with surfaces Ω_{ij}, $i, j = 1, \ldots, m$,

$\Omega = \bigcup_{i,j=1}^{m} \Omega_{ij}$, the WFS measurements can be modeled as an operator equation

$$\mathbf{s} = \Gamma \varphi, \tag{1.8}$$

where $\Gamma := [\Gamma_{\mathbf{x}}, \Gamma_{\mathbf{y}}]^T$, and

$$\begin{aligned} \mathbf{s}_{\mathbf{x}}[i,j] = \Gamma_{\mathbf{x}}(\varphi)[i,j] &:= \frac{1}{|\Omega_{ij}|} \int_{\Omega_{ij}} \frac{\partial \varphi}{\partial \mathbf{x}}(\mathbf{x},\mathbf{y})\, d(\mathbf{x},\mathbf{y}), \\ \mathbf{s}_{\mathbf{y}}[i,j] = \Gamma_{\mathbf{y}}(\varphi)[i,j] &:= \frac{1}{|\Omega_{ij}|} \int_{\Omega_{ij}} \frac{\partial \varphi}{\partial \mathbf{y}}(\mathbf{x},\mathbf{y})\, d(\mathbf{x},\mathbf{y}). \end{aligned} \tag{1.9}$$

The operator Γ is well defined for wavefronts $\varphi \in H^s$, $s > 1/2$, as shownin [31]. With wavefronts following the van Karman power law (1.6), we have:

$$\|C_\varphi^{-1/2}\varphi\|_{L^2}^2 \simeq \|\varphi\|_{H^{11/6}}$$

should be bounded. Therefore, it makes sense to consider Γ as

$$\Gamma : H^{11/6}(\Omega) \to \mathbb{R}^{m\times m\times 2}, \tag{1.10}$$

mapping the phase φ onto measurements $\mathbf{s}$ [25, 45, 52, 32]. We give the discretized version of (1.9) in the following context.

Discretized model. Assume Ω is a square, Ω_{ij} are sub-sqaures of Ω with equal area, and there are r^2 discretization points in each direction in Ω_{ij}, denoted by Ω_{ij}^{ks} with $1 \le k, s \le r$. That is, $\Omega_{ij} = \cup_{k,s=1}^{r} \Omega_{ij}^{ks}$. For any $(\mathbf{x},\mathbf{y}) \in \Omega_{ij}^{ks}$, we assume $\varphi(\mathbf{x},\mathbf{y}) = \varphi_{ij}^{ks}$ and approximate $\nabla\varphi(\mathbf{x},\mathbf{y})$ by

$$\begin{aligned} \frac{\partial \varphi}{\partial \mathbf{x}}(\mathbf{x},\mathbf{y}) &= \frac{(\varphi_{ij}^{k+1,s} - \varphi_{ij}^{ks}) + (\varphi_{ij}^{k+1,s+1} - \varphi_{ij}^{k,s+1})}{2}, \\ \frac{\partial \varphi}{\partial \mathbf{y}}(\mathbf{x},\mathbf{y}) &= \frac{(\varphi_{ij}^{k,s+1} - \varphi_{ij}^{ks}) + (\varphi_{ij}^{k+1,s+1} - \varphi_{ij}^{k+1,s})}{2}. \end{aligned} \tag{1.11}$$

Here, we ignore the length of the edge in Ω_{ij}^{ks} by assuming it equals 1. Note that as the edge length goes to 0, the above approximation converges to the true model in (1.9). Using the periodic boundary condition, the model (1.9) can be calculated by

$$\begin{aligned} \mathbf{s}_{\mathbf{x}}[i,j] &= \frac{1}{r^2} \sum_{k,s=1}^{r} \frac{(\varphi_{ij}^{k+1,s} - \varphi_{ij}^{ks}) + (\varphi_{ij}^{k+1,s+1} - \varphi_{ij}^{k,s+1})}{2}, \\ \mathbf{s}_{\mathbf{y}}[i,j] &= \frac{1}{r^2} \sum_{k,s=1}^{r} \frac{(\varphi_{ij}^{k,s+1} - \varphi_{ij}^{ks}) + (\varphi_{ij}^{k+1,s+1} - \varphi_{ij}^{k+1,s})}{2}. \end{aligned} \tag{1.12}$$

Defining $n = rm$ and discretizing φ as $\boldsymbol{\phi} \in \mathbb{R}^{n\times n}$, we can rearrange the index in (1.12) and obtain

$$\mathbf{s}_{\mathbf{x}} = \Gamma_x^d(\boldsymbol{\phi}) = \downarrow \circ D_{\mathbf{x}}(\boldsymbol{\phi}), \quad \mathbf{s}_{\mathbf{y}} = \Gamma_y^d(\boldsymbol{\phi}) = \downarrow \circ D_{\mathbf{y}}(\boldsymbol{\phi}), \tag{1.13}$$

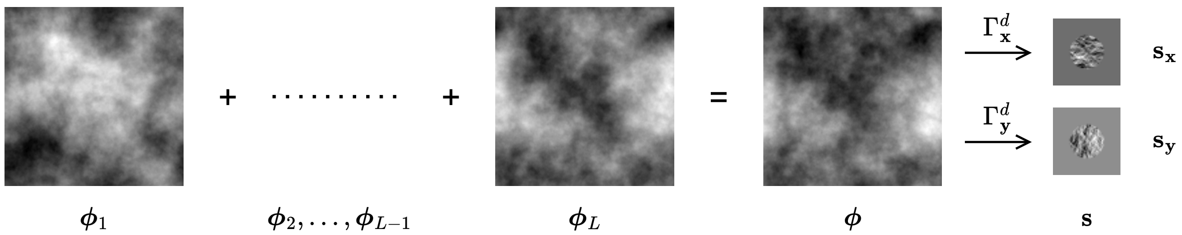


Figure 3: The incoming wavefront $\boldsymbol{\phi}$ equals to the sum of multi-layer phases $\boldsymbol{\phi}_1$ to $\boldsymbol{\phi}_L$, which can be observed through the wavefront sensor $\Gamma^d_{\mathbf{x}}$ and $\Gamma^d_{\mathbf{y}}$.

where

$$\downarrow(\boldsymbol{\psi})[s,t] := \frac{1}{r^2}\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\boldsymbol{\psi}[r(s-1)+i, r(t-1)+j], \tag{1.14}$$

and

$$\begin{aligned} D_{\mathbf{x}}(\boldsymbol{\phi})[k,l] &:= \frac{(\boldsymbol{\phi}[k+1,l]-\boldsymbol{\phi}[k,l])+(\boldsymbol{\phi}[k+1,l+1]-\boldsymbol{\phi}[k,l+1])}{2}, \\ D_{\mathbf{y}}(\boldsymbol{\phi})[k,l] &:= \frac{(\boldsymbol{\phi}[k,l+1]-\boldsymbol{\phi}[k,l])+(\boldsymbol{\phi}[k+1,l+1]-\boldsymbol{\phi}[k+1,l])}{2}. \end{aligned} \tag{1.15}$$

Figure 3 illustrates this forward process. The goal of this work is to reconstruct the high-resolution phase $\boldsymbol{\phi} \in \mathbb{R}^{n\times n}$ through the low-resolution phase gradient $\mathbf{s} \in \mathbb{R}^{m\times m\times 2}$. However, when $r > 1$, the number of unknowns is $n^2 = r^2m^2$, which exceeds the number of observations $2m^2$. Thus, it is difficult to recover $\boldsymbol{\phi}$ from $\mathbf{s}$ directly. To address this, we consider the K multi-frame wavefront gradient observations during a short time interval, i.e.

$$\mathbf{s}^i_{\mathbf{x}} = \Gamma^d_x(\boldsymbol{\phi}^i), \quad \mathbf{s}^i_{\mathbf{y}} = \Gamma^d_y(\boldsymbol{\phi}^i), \quad i = 1, 2, \cdots, K. \tag{1.16}$$

where $\boldsymbol{\phi}^i \in \mathbb{R}^{n\times n}$ and K is the number of time points. The following context simplifies (1.16) by relating $\boldsymbol{\phi}^i$ and $\boldsymbol{\phi}^j$ under the Taylor Frozen Flow (TFF) hypothesis.

The TFF hypothesis, introduced in [43], claims that the atmosphere consists of layers of distinguishable turbulence that move within short intervals parallel to the Earth's surface at a certain velocity. In other words, the turbulence pattern does not change within those small time intervals but only moves in a certain direction with a certain velocity. Furthermore, we assume each layer's atmosphere is of linear motion with a constant velocity. Mathematically, assuming there are L layers and the i-th frame with wind shift between two consecutive frames in layer l is v_l for $1 \le l \le K$, we have

$$\boldsymbol{\phi} = \sum_{l=1}^{L}\boldsymbol{\phi}_l \quad \text{and} \quad \boldsymbol{\phi}^i = \sum_{l=1}^{L}\boldsymbol{\phi}_l(\cdot - (i-1)v_l).$$

Here, $\boldsymbol{\phi}_l(\cdot - v_l)$ means moving the matrix $\boldsymbol{\phi}_l$ linearly along the v_l direction, which can be implemented using bilinear interpolation. In practice, the velocity $\{v_l\}_{l=1}^{L}$ of the atmosphere can be measured experimentally by sending a balloon into the atmosphere. Future telescopes might even be equipped with instruments that allow us to measure the wind speed in the atmosphere directly. Define the evolution operator as $A_l^i \phi_l(\cdot) = \phi_l(\cdot - (i-1)v_l)$ and under the TFF hypothesis, the forward model in our phase reconstruction problem with multi-frame observation is

$$\mathbf{s}^i := \begin{bmatrix} \mathbf{s}_{\mathbf{x}}^i \\ \mathbf{s}_{\mathbf{y}}^i \end{bmatrix} = \begin{bmatrix} \Gamma_{\mathbf{x}}^d \\ \Gamma_{\mathbf{y}}^d \end{bmatrix} \left(\sum_{l=1}^{L} A_l^i \phi_l \right) + \begin{bmatrix} \mathbf{n}_{\mathbf{x}}^i \\ \mathbf{n}_{\mathbf{y}}^i \end{bmatrix}, \quad i = 1, \ldots, K, \tag{1.17}$$

where $\mathbf{n}_{\mathbf{x}}^i, \mathbf{n}_{\mathbf{y}}^i$ are noise. Our goal is to reconstruct the high-resolution incoming wavefront $\phi = \sum_{l=1}^{L} \phi_l \in \mathbb{R}^{n \times n}$ from multi-frame low-resolution wavefront gradient $\{\mathbf{s}^i \in \mathbb{R}^{m \times m \times 2}\}_{i=1}^{K}$. Our main contributions in the paper are summarized as follows.

- To solve (1.17), we propose a variational model that contains the traditional Laplacian regularization and a data-driven regularization. Utilizing training samples, the proposed new regularization term can be deduced by resolving a bilevel optimization problem that compensates for the approximation error between traditional regularization and the underlying true distribution of the signals.
- We represent the gradient of the data-driven regularization by a deep neural network and unroll the Nesterov accelerated gradient algorithm to minimize the inner problem, leading to the so-called PhaseNet. Different from other unrolling methods, the network in each layer shares the same parameters, which significantly reduces the network size and can obtain an unrolling network with many layers.
- Experimental results on wavefront reconstruction with different atmospheric conditions validate the advantages of the proposed PhaseNet over traditional variational methods and unrolling approaches.

The rest of the paper is organized as follows. In Section 2, we discuss related works of phase reconstruction in ground-based astronomy, including the phase gradient model, phase model, and deep learning based model. In Section 3, we propose our energy function with a data-driven residual regularization term and introduce the phase reconstruction network PhaseNet. In Section 4, we compare our PhaseNet with TV and Laplacian models on various atmospheric conditions and analyze the performance of the proposed method from six perceptions. The conclusion is given in Section 5.

2. Related work.

This section briefly reviews the work closely related to phase reconstruction in ground-based astronomy, including the phase gradient, phase, and deep learning models.

Phase gradient model. To estimate the incoming wavefront phase ϕ, one idea is to decompose the WFS operator into the derivative operator $D = [D_{\mathbf{x}}, D_{\mathbf{y}}]^T$ and the downsampling operator $\downarrow$, *i.e.*

$$\Gamma^d = \downarrow \circ D, \tag{2.1}$$

and reconstruct ϕ through a two-stage process. The first stage is to reconstruct the high-resolution wavefront gradient $\phi_{\mathbf{s}}$ by solving an ill-posed inverse problem:

$$\mathbf{s} = \downarrow \phi_{\mathbf{s}} + \mathbf{n}, \tag{2.2}$$

which is an image super-resolution problem in low-level vision. Then the phase $\boldsymbol{\phi}$ can be recovered by integrating from $\boldsymbol{\phi}_{\mathbf{s}}$, *i.e.*

$$\boldsymbol{\phi} = D\boldsymbol{\phi}_{\mathbf{s}}, \tag{2.3}$$

which is an over-determined linear system [13, 15, 35, 42]. However, since (2.2) is an under-determined problem, it is difficult to obtain a high-quality wavefront gradient $\boldsymbol{\phi}_{\mathbf{s}}$. To address this dilemma, Jefferies and Hart [20] proposed to use the multi-frame observation to recover $\boldsymbol{\phi}_{\mathbf{s}}$ under the TFF hypothesis [29]. Moreover, this multi-frame method is improved in [10, 5] with Tikhonov and $l^1 - l^p$ regularizations.

Phase model. Unlike the two-stage phase gradient model, the phase model chooses to solve (1.17) directly with different regularization terms [9, 6, 50, 22, 2] on the phase $\boldsymbol{\phi}$ such as total variation, Laplacian, and Huber norm. It is shown in [6] that the reconstruction performance of the phase model is better than the phase gradient model. One possible reason is that the phase gradient model needs to solve two sub-problems, and the error of the first sub-problem will be amplified when solving the second one. Moreover, it is easier to design a regularization for phase $\boldsymbol{\phi}$ rather than its gradient $\boldsymbol{\phi}_{\mathbf{s}}$ since the phase behaves more like natural images while the phase gradient is more ambiguous.

Deep learning model. Deep learning methods have shown great potential in solving mathematical inverse problems in recent years. In ground-based astronomy, most of the existing works use Multilayer Perceptrons (MLPs) or Convolutional Neural Networks (CNN) to estimate the Zernike coefficients, that is, the coefficients of the representation

$$\varphi(\mathbf{x}, \mathbf{y}) = \sum_{n,m} a_{nm} Z_n^m(\mathbf{x}, \mathbf{y}) \tag{2.4}$$

of the high-resolution incoming wavefront from the noisy wavefront gradient data or the whole off-axis SH-WFS images [17, 27, 11, 18, 19]. Here, Z_n^m are the Zernike polynomials, and a_{nm} are the Zernike coefficients [33], (m, n) is an indexing scheme with n being the radial order and m the angular order. In addition, Swanson et al. [41] proposed a U-Net [39] based neural network to predict the wavefront image from low-resolution wavefront gradient observations. However, these methods use the neural network as a black box solver without considering the underlying mathematical model. Despite the satisfactory numerical performance achieved by deep learning based methods, the lack of interpretability remains a significant concern when comparing them with model-based approaches. The unrolling method is an emerging technique that alleviates the interpretability issues in signal and image processing [40, 28] and is becoming increasingly popular. This technique forms a deep neural network by unrolling a traditional iterative numerical algorithm and replacing the map in each iteration with a single network. Compared to traditional deep learning methods, the unrolling approach is motivated by solving a variational model and provides more insights into network design. In this work, we apply the unrolling idea to solve the phase reconstruction task in ground-based astronomy. We directly model the gradient of the data-driven regularization term using a single network, enhancing our method's explainability. Moreover, we utilize multi-frame observations as the input, further improving the reconstruction accuracy.

3. Our methodology. Our goal is to recover the incoming phase ϕ from the multi-frame low-resolution wavefront gradient $\{\mathbf{s}^i\}_{i=1}^K$ with a neural network $\mathbf{F}_\theta$, where θ denotes the network parameter. The neural network $\mathbf{F}_\theta$ can be trained by minimizing a designed loss given a set of training samples. Recall that the forward model (1.17) can be written as

$$\begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \vdots \\ \mathbf{s}^K \end{bmatrix} = \begin{bmatrix} \Gamma^d A_1^1 & \Gamma^d A_2^1 & \cdots & \Gamma^d A_L^1 \\ \Gamma^d A_1^2 & \Gamma^d A_2^2 & \cdots & \Gamma^d A_L^2 \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma^d A_1^K & \Gamma^d A_2^K & \cdots & \Gamma^d A_L^K \end{bmatrix} \begin{bmatrix} \phi^1 \\ \phi^2 \\ \vdots \\ \phi^L \end{bmatrix} + \begin{bmatrix} \mathbf{n}^1 \\ \mathbf{n}^2 \\ \vdots \\ \mathbf{n}^K \end{bmatrix}, \tag{3.1}$$

where $\Gamma^d = [\Gamma_x^d, \Gamma_y^d]^\top$. For simplicity, we rewrite (3.1) as

$$\bar{\mathbf{s}} = \mathbf{H}\mathbf{\Phi} + \mathbf{n} \tag{3.2}$$

where $\mathbf{H}$ is the linear operator in (3.1), $\bar{\mathbf{s}} = [\mathbf{s}^1, \ldots, \mathbf{s}^K]^T$ denotes all the wavefront gradient frames, $\mathbf{\Phi} = [\phi_1, \ldots, \phi_L]^T$ indicates the atmospheric turbulence presence in each layer.

The classical method to recover the phase $\mathbf{\Phi}$ is solving the following optimization problem:

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\mathbf{R}(\mathbf{\Phi}), \tag{3.3}$$

where $\mathbf{R}$ denotes a regularization for $\mathbf{\Phi}$. The final incoming wavefront ϕ can be obtained as the summation of ϕ_l, $l = 1, \ldots, L$. To identify an appropriate regularization term, it is necessary to reference turbulence statistics. According to Kolmogorov's theory [24, 47], atmospheric turbulence within each layer is assumed as a homogeneous and isotropic Gaussian process with a covariance operator C_φ as outlined in (1.4). Concurrently, our application of the von Karman power spectral density model in (1.6) indicates that the regularization for each layer phase φ might be chosen:

$$\mathbf{R}(\varphi) = \|C_\varphi^{-1/2}\varphi\|_{L^2}^2 = C_0\|(\kappa_0^2 + |\kappa|^2)^{\frac{11}{12}}\mathcal{F}\varphi\|_{L^2}^2 \simeq \kappa_0^{\frac{11}{3}}\|\varphi\|_{L^2}^2 + \|(-\Delta)^{\frac{11}{12}}\varphi\|_{L^2}^2, \tag{3.4}$$

where C_0 is a constant. Note that $\|C_\varphi^{-1/2}\varphi\|_{L^2}^2$ is bounded as we assume $\varphi \in H^{11/6}$. In the discretized model, the covariance operator for the phase ϕ_l in l-th layer becomes a covariance matrix denoted by C_{ϕ_l}. Standard reconstruction algorithms use $\|C_{\phi_l}^{-1/2}\phi_l\|_2$ as the regularization function [22, 14]. However, in practice, the covariance matrix C_{ϕ_l} is a dense matrix, making its application computationally inefficient for large-scale problems. Many approximation methods have been developed in recent decades to simplify the covariance matrix [44, 49]. From the equivalence in (3.4), Ellerbroek [13] proposed to use the biharmonic operator Δ^2 to approximate the inverse covariance operator C_φ^{-1} [13]. Furthermore, we assume atmospheric turbulence layers are mutually independent, rendering the regularization function as

$$\mathbf{R}(\mathbf{\Phi}) = \sum_{l=1}^{L}\mathbf{R}(\phi_l) = \sum_{l=1}^{L}\|\mathbf{L}\phi_l\|_2^2, \tag{3.5}$$

where $\mathbf{L}$ represents the discrete Laplacian operator.

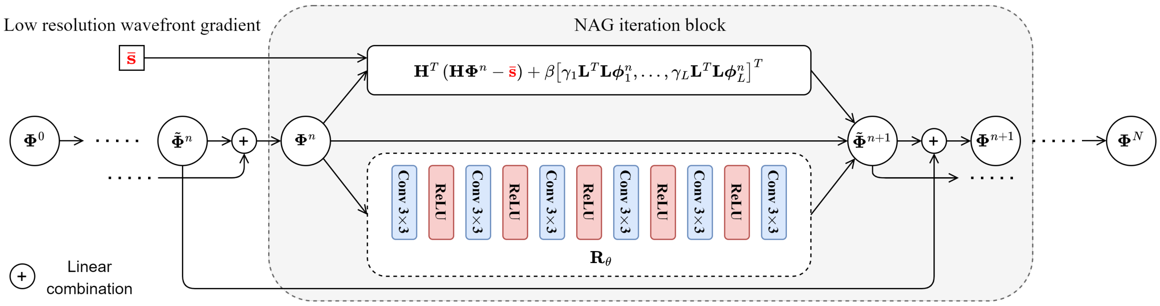


Figure 4: Data flow and network architecture in PhaseNet.

Algorithm 3.1 Forward propagation of PhaseNet

Input: Multi-frame wavefront gradient $\bar{\mathbf{s}}$, $\mathbf{\Phi}^0$, step sizes $\{\alpha_n\}_{n=1}^N$, number of iterations N.
Output: Reconstructed incoming wavefront $\boldsymbol{\phi} = \sum_{j=1}^L \boldsymbol{\phi}_j$.

Initialize $\lambda_0 = 1$, $\tilde{\mathbf{\Phi}}^0 = \mathbf{\Phi}^0$;
for $n = 0, 1, 2, \ldots, N$ **do**
  $\lambda_{n+1} = \frac{1+\sqrt{1+4\lambda_n^2}}{2}$, $\eta_n = \frac{1-\lambda_n}{\lambda_{n+1}}$;
  $\tilde{\mathbf{\Phi}}^{n+1} = \mathbf{\Phi}^n - \alpha_n \Big(\mathbf{H}^T(\mathbf{H}\mathbf{\Phi}^n - \bar{\mathbf{s}}) + \beta\left[\gamma_1\mathbf{L}^T\mathbf{L}\boldsymbol{\phi}_1^n, \ldots, \gamma_L\mathbf{L}^T\mathbf{L}\boldsymbol{\phi}_L^n\right]^T + \mathbf{R}_\theta(\mathbf{\Phi}^n)\Big)$;
  $\mathbf{\Phi}^{n+1} = (1-\eta_n)\tilde{\mathbf{\Phi}}^{n+1} + \eta_n\tilde{\mathbf{\Phi}}^n$;
end for
return Incoming wavefront $\boldsymbol{\phi} = \sum_{l=1}^L \boldsymbol{\phi}_l^{N+1}$.

In this case, the optimization problem in (3.3) becomes the Laplacian regularized model:

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^L \gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2, \tag{3.6}$$

where γ_l denotes the layer weight. However, the Laplacian regularization is imperfect and different from the actual regularization function due to the accumulation of several approximation errors, such as the difference between the biharmonic operator Δ^2 and inverse covariance operator $C_{\boldsymbol{\phi}}^{-1}$ and the error in Kolmogorov's turbulence statistics. To alleviate the problem, we propose to add a residual term $\mathbf{R}_\vartheta^{res}(\cdot)$ to represent the difference between Laplacian regularization and the underlying real regularization, where ϑ denotes the learnable parameters. In summary, our model is given as

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^L \gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2 + \mathbf{R}_\vartheta^{res}(\mathbf{\Phi}), \tag{3.7}$$

where $\beta > 0$ is the penalty parameter.

Obtaining the precise formulation of the residual regularization $\mathbf{R}_\vartheta^{res}(\mathbf{\Phi})$ is challenging. Consequently, we employ a data-driven approach that implicitly learns the residual regularization. Specifically, we consider training samples represented by $\{\bar{\mathbf{s}}^n, \boldsymbol{\phi}_{\text{true}}^n\}_{n=1}^{\mathbf{N}}$, where $\bar{\mathbf{s}}^n$

signifies the multi-frame wavefront gradient, ϕ embodies the high-resolution incident wavefront, and $\mathbf{N}$ indicates the quantity of training instances. Our objective is to ascertain the residual regularization function $\mathbf{R}_{\vartheta}^{res}(\mathbf{\Phi})$ by solving the following bilevel optimization problem:

$$
\begin{aligned}
&\min_{\vartheta} \sum_{n=1}^{\mathbf{N}} \ell(\sum_{l=1}^{L} \phi_{l\star}^{n}, \phi_{true}^{n}), \quad \text{where } \mathbf{\Phi}_{\star}^{n} = [\phi_{1\star}^{n}, \dots, \phi_{L\star}^{n}]^{T} \\
&\text{s.t. } \mathbf{\Phi}_{\star}^{n} = \arg\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}}^{n} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^{L}\gamma_l\|\mathbf{L}\phi_l\|_2^2 + \mathbf{R}_{\vartheta}^{res}(\mathbf{\Phi}),
\end{aligned} \tag{3.8}
$$

where $\ell(\cdot,\cdot)$ denotes the loss function. Since solving the above bilevel optimization is difficult, we mimic the minimization of the inner problem by an unrolled deep neural network and reduce (3.8) to a single-level minimization problem.

The basic idea comes from the gradient descent algorithm that solves the inner optimization problem in (3.8). Each update has the form:

$$
\mathbf{\Phi}^{n+1} = \mathbf{\Phi}^{n} - \alpha_n \left(\mathbf{H}^T(\mathbf{H}\mathbf{\Phi}^n - \bar{\mathbf{s}}) + \beta\left[\gamma_1\mathbf{L}^T\mathbf{L}\phi_1^n, \dots, \gamma_L\mathbf{L}^T\mathbf{L}\phi_L^n\right]^T + \nabla\mathbf{R}_{\vartheta}^{res}(\mathbf{\Phi}^n)\right), \tag{3.9}
$$

where α_n denotes the step size in the n-th iteration. From the gradient descent iteration, we find that we only need $\mathbf{R}_{\vartheta}^{res}$ to solve the lower-level optimization problem; therefore, instead of learning the residual regularization $\mathbf{R}_{\vartheta}^{res}(\mathbf{\Phi})$, we learn its gradient. In particular, we use a neural network to parameterize the gradient of $\mathbf{R}_{\vartheta}^{res}$, *i.e.*

$$
\mathbf{R}_{\theta}(\cdot) = \nabla\mathbf{R}_{\vartheta}^{res}(\cdot), \tag{3.10}
$$

where $\mathbf{R}_{\theta}(\cdot)$ is a neural network with parameter θ. Specifically, $\mathbf{R}_{\theta}$ is a Convolutional Neural Network (CNN) consisting of six 3×3 convolution layers with 64 channels and Rectified Linear Unit (ReLU) activation functions. Combining (3.10) with (3.9), the gradient descent iteration becomes:

$$
\mathbf{\Phi}^{n+1} = \mathbf{\Phi}^{n} - \alpha_n \left(\mathbf{H}^T(\mathbf{H}\mathbf{\Phi}^n - \bar{\mathbf{s}}) + \beta\left[\gamma_1\mathbf{L}^T\mathbf{L}\phi_1^n, \dots, \gamma_L\mathbf{L}^T\mathbf{L}\phi_L^n\right]^T + \mathbf{R}_{\theta}(\mathbf{\Phi}^n)\right), \tag{3.11}
$$

and the entire iterative process can be unrolled as a trainable neural network. Due to the slow convergence of gradient descent, we adopt an extrapolation method to accelerate the optimization process. The modified iteration is then given by:

$$
\begin{aligned}
\tilde{\mathbf{\Phi}}^{n+1} &= \mathbf{\Phi}^{n} - \alpha_n \left(\mathbf{H}^T(\mathbf{H}\mathbf{\Phi}^n - \bar{\mathbf{s}}) + \beta\left[\gamma_1\mathbf{L}^T\mathbf{L}\phi_1^n, \dots, \gamma_L\mathbf{L}^T\mathbf{L}\phi_L^n\right]^T + \mathbf{R}_{\theta}(\mathbf{\Phi}^n)\right), \\
\mathbf{\Phi}^{n+1} &= (1-\eta_n)\tilde{\mathbf{\Phi}}^{n+1} + \eta_n\tilde{\mathbf{\Phi}}^{n},
\end{aligned} \tag{3.12}
$$

where η_n denotes the extrapolation factor. In practice, the Nesterov Accelerated Gradient (NAG) method [30] is employed to determine η_n. The detailed forward process of the NAG method is outlined in Algorithm 3.1, while the final unrolled network, referred to as the PhaseNet, comprises a series of NAG iteration blocks, as illustrated in Figure 4. A comparison of different optimization algorithms is provided in Table 6. It is worth mentioning that the NAG iteration blocks share the same parameters, resulting in a smaller network size and better consistency with respect to minimizing the inner problem (3.8).

Table 1: Simulation parameters.

Telescope diameter d	8m
WFS wavelength λ	744 nm
Resolution incoming wavefront	200×200
Resolution wavefront gradient	50×50

Remark 3.1. The convergence of the iteration given in (3.12) can be obtained under certain assumptions and restrictions on $\mathbf{R}_\theta$ and η_n, as shown in [48, 1, 4].

Given that PhaseNet can effectively address the lower-level optimization task, the bilevel optimization problem in (3.8) can be simplified to

$$\min_{\theta} \sum_{n=1}^{\mathbf{N}} \ell(\text{PhaseNet}_\theta(\bar{\mathbf{s}}^n), \boldsymbol{\phi}_{true}^n), \tag{3.13}$$

which is optimized with deep learning training strategies. After the training phase, for an emergent wavefront gradient $\mathbf{s}$, the high-resolution incident wavefront phase $\boldsymbol{\phi}$ can be reconstructed through PhaseNet's evaluation, mimicking to minimize (3.7).

Since our goal is to reconstruct the PSF in (1.1), we choose the relative error between the reconstructed PSF with ground truth PSF as the loss function of our PhaseNet. Assume $\boldsymbol{\phi}$ is the estimated incoming wavefront by our PhaseNet, $\boldsymbol{\phi}_{true}$ is the ground truth wavefront, the loss function is

$$\ell(\boldsymbol{\phi}, \boldsymbol{\phi}_{true}) = \text{Relative Error}(\mathbf{k}, \mathbf{k}_{true}) := \frac{\|\mathbf{k} - \mathbf{k}_{true}\|_2}{\|\mathbf{k}_{true}\|_2} \tag{3.14}$$

where $\mathbf{k} = |\mathcal{F}^{-1}\{\mathcal{P}\exp[\iota\boldsymbol{\phi}]\}|^2$, and $\mathbf{k}_{true} = |\mathcal{F}^{-1}\{\mathcal{P}\exp[\iota\boldsymbol{\phi}_{true}]\}|^2$.

Remark 3.2. In PhaseNet, the hyper-parameters, including the step sizes $\{\alpha_n\}_{n=1}^N$ and $\{\eta_n\}_{n=1}^N$ used in Algorithm 3.1, have been configured as learnable parameters, leading to an improvement in performance.

4. Experiments and results.

In this section, we evaluate the performance of PhaseNet with different turbulence atmosphere conditions and compare our network with traditional variational based methods. To accomplish this, we obtained the training and testing data using a simulation tool based on MATLAB [2]. The telescope parameters utilized in the simulation are outlined in Table 1, wherein an 8m telescope is employed, equipped with a single 50×50 SH-WFS.

4.1. Implementation details.

In the forward model of our PhaseNet, we employ 1,000 NAG steps. The training process of all PhaseNet models consists of 30,000 iterations with Adam optimiser [23]. The initial learning rate is initialized to 1×10^{-4} and is subsequently reduced by half every 10,000 iterations. The batch size is set to 1. The initial step size $\{\alpha_n\}_{n=1}^N$ is set to 0.5, while β is fixed at 1×10^{-4}, and γ_l is set to 1 for each layer. The network architecture is illustrated in Figure 4. In contrast to conventional deep unrolling

Table 2: Comparison of averaged relative errors of phase with seeing condition $d/r_0 = 10$.

Dataset	TV model	Laplacian model	PhaseNet (ours)
Two-frames-one-layer	0.0331	0.0303	**0.0302**
Four-frames-one-layer	0.0220	0.0212	**0.0203**
Eight-frames-one-layer	0.0153	0.0156	**0.0142**
Two-frames-three-layers	0.0483	0.0454	**0.0439**
Four-frames-three-layers	0.0410	0.0383	**0.0381**

techniques, we employ numerous iterations to ensure the precise reconstruction of PSF in our PhaseNet. However, this requirement entails a considerable demand for GPU memory. To address this problem, we only employ the network to correct the Laplacian gradient every 20 steps.

4.2. Datasets and evaluation metric. To evaluate the effectiveness of our PhaseNet model in reconstructing incoming wavefronts in ground-based astronomy, we simulated five datasets under varying atmospheric conditions. Each dataset is named based on the number of subsequent observation frames and atmosphere layers used in the simulation. The observation frames in one data sample are coming from a single WFS.

- **Two frames one layer.** We use the method in [2] to obtain 1,000 paired $\mathbf{\Phi}$ and $\bar{\mathbf{s}}$ as the training dataset. We assume the wind's direction in each turbulence layer is uniformly generated on a unit circle S^1, and the wind speed is bounded by 10 pixels per frame. The number of turbulence layers is 1, and there are 2 frames for the wavefront gradient $\bar{\mathbf{s}}$ with 1% Gaussian noise. In the training stage, we assume the seeing condition d/r_0 changes from 5 to 45, where a smaller seeing condition means a better observation environment in ground-based astronomy. So, our network can be applied to different atmospheric conditions. In the testing stage, we first generate 20 samples with the seeing condition being 10 and then generate another 20 test samples with the seeing condition being 40 to evaluate our model. We refer to this dataset as the Two-frames-one-layer dataset.
- **Four frames one layer.** The setting is the same as the Two-frames-one-layer dataset, except there are 4 frames in one data sample. This dataset is referred as the Four-frames-one-layer dataset.
- **Eight frames one layer.** The setting is the same as the Two-frames-one-layer dataset, except there are 8 frames in one data sample, and the wind speed is bounded by 5 pixel-per-frame. This dataset is referred as the Eight-frames-one-layer dataset.
- **Two frames three layers.** The setting is the same as the Two-frames-one-layer dataset, except there are 3 turbulence layers. The heights of each layer are 0m, 11,000m, and 15,000m. This dataset is referred to as the Two-frames-three-layers dataset.
- **Four frames three layers.** The setting is the same as the Two-frames-three-layer dataset, except there are 4 frames in one data sample. This dataset is referred to as the Four-frames-three-layers dataset.

Table 3: Comparison of averaged relative errors of phase with seeing condition $d/r_0 = 40$.

Dataset	TV model	Laplacian model	PhaseNet (ours)
Two-frames-one-layer	0.0318	0.0291	**0.0288**
Four-frames-one-layer	0.0212	0.0204	**0.0193**
Eight-frames-one-layer	0.0147	0.0150	**0.0133**
Two-frames-three-layers	0.0470	0.0444	**0.0429**
Four-frames-three-layers	0.0398	0.0375	**0.0339**

Table 4: Comparison of averaged relative errors of PSF with seeing condition $d/r_0 = 10$.

Dataset	TV model	Laplacian model	PhaseNet (ours)
Two-frames-one-layer	0.0489	0.0293	**0.0250**
Four-frames-one-layer	0.0203	0.0158	**0.0116**
Eight-frames-one-layer	0.0113	0.0098	**0.0068**
Two-frames-three-layers	0.0863	0.0702	**0.0498**
Four-frames-three-layers	0.0672	0.0568	**0.0384**

Evaluation metric. We present the averaged relative error of the estimated phase as the quantitative results of our method, which is defined as

$$\text{Phase Relative Error} = \frac{\|\boldsymbol{\phi} - \boldsymbol{\phi}_{true}\|_2}{\|\boldsymbol{\phi}_{true}\|_2}, \tag{4.1}$$

where $\boldsymbol{\phi}_{true}$ represents the ground truth phase, while $\boldsymbol{\phi}$ denotes the estimated phase.

Remark 4.1. As the WFS employed in our model cannot distinguish between two phases up to a constant shift, the restored phase inevitably possesses an unknown constant shift compared to the ground truth phase. To rectify this bias, we normalize the estimated phase $\boldsymbol{\phi}$ and the ground truth phase $\boldsymbol{\phi}_{true}$ by setting their means to zero before computing the phase relative error. Specifically, we use the following normalization procedure:

$$\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} - \frac{\sum_{ij} \boldsymbol{\phi}[i,j]}{n^2}, \quad \boldsymbol{\phi}_{true} \leftarrow \boldsymbol{\phi}_{true} - \frac{\sum_{ij} \boldsymbol{\phi}_{true}[i,j]}{n^2}, \tag{4.2}$$

where n represents the spatial size of the phase.

Moreover, since our goal is to recover the PSF and thereby restore the blurred observation, we compute the relative error of the estimated PSF as an evaluation metric, which is defined as

$$\text{PSF Relative Error} = \frac{\|\mathbf{k} - \mathbf{k}_{true}\|_2}{\|\mathbf{k}_{true}\|_2}, \tag{4.3}$$

where $\mathbf{k} = |\mathcal{F}^{-1}\{\mathcal{P}\exp[\iota\boldsymbol{\phi}]\}|^2$, and $\mathbf{k}_{true} = |\mathcal{F}^{-1}\{\mathcal{P}\exp[\iota\boldsymbol{\phi}_{true}]\}|^2$.

Table 5: Comparison of averaged relative errors of PSF with seeing condition $d/r_0 = 40$.

Dataset	TV model	Laplacian model	PhaseNet (ours)
Two-frames-one-layer	0.3007	0.2638	**0.2459**
Four-frames-one-layer	0.1810	0.1733	**0.1542**
Eight-frames-one-layer	0.1187	0.1216	**0.1023**
Two-frames-three-layers	0.4778	0.4455	**0.4036**
Four-frames-three-layers	0.3962	0.3699	**0.3203**

4.3. Results. We compare our PhaseNet with two traditional variational methods, namely the TV model [6], and the Laplacian model [22]. The objective function of the TV model is

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H\Phi}\|_2^2 + \beta \sum_{l=1}^{L} \gamma_l \|\nabla \boldsymbol{\phi}_l\|_1, \tag{4.4}$$

which is solved through the Alternating Direction Method of Multipliers (ADMM) algorithm. The objective function for the Laplacian model is:

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H\Phi}\|_2^2 + \frac{\beta}{2} \sum_{l=1}^{L} \gamma_l \|\Delta \boldsymbol{\phi}_l\|_2^2, \tag{4.5}$$

which can be solved through the Conjugate Gradient (CG) method.

Comparison of Phase. We present the averaged relative errors of the reconstructed phase for all five datasets, with seeing conditions of 10 and 40, in Tables 2 and 3, respectively. The results demonstrate that the proposed PhaseNet outperforms traditional variational methods, achieving the lowest phase relative error. Additionally, in Figure 5, we provide visual representations of the error image for the reconstructed phase. These images demonstrate that our method recovers more high-frequency information than TV and Laplacian methods.

Comparison of PSF. As our goal is to recover the PSF, we present the PSF relative errors on five testing datasets in Figure 6. The figure shows that PhaseNet achieves the lowest relative error in PSF reconstruction on almost all testing samples compared to TV and Laplacian models. The average relative errors for seeing conditions of 10 and 40 are presented in Table 4 and Table 5, respectively, with our PhaseNet outperforming traditional methods on all datasets and atmospheric conditions. The PSF relative error of PhaseNet is improved by 0.02 and 0.01 on average compared with TV and Laplacian models on five datasets with seeing condition 10 and 0.05 and 0.03 with seeing condition 40. Additionally, Figure 10–Figure 13 (first row) shows the visualizations of the estimated PSFs. From the enlarged area in these Figures, we find that the reconstructed PSFs using our method or the ground truth fit more closely than traditional TV and Laplacian models.

Deconvolution Results. We use the estimated PSF to deblur the observation, with results shown in Figure 10–Figure 13 (second and third rows). The deconvolution problem is solved through the ADMM algorithm using the objective function

$$\min_{f} \frac{1}{2}\|g(\mathbf{x},\mathbf{y}) - (k * f)(\mathbf{x},\mathbf{y})\|_2^2 + \beta\|\nabla f(\mathbf{x},\mathbf{y})\|_1, \tag{4.6}$$

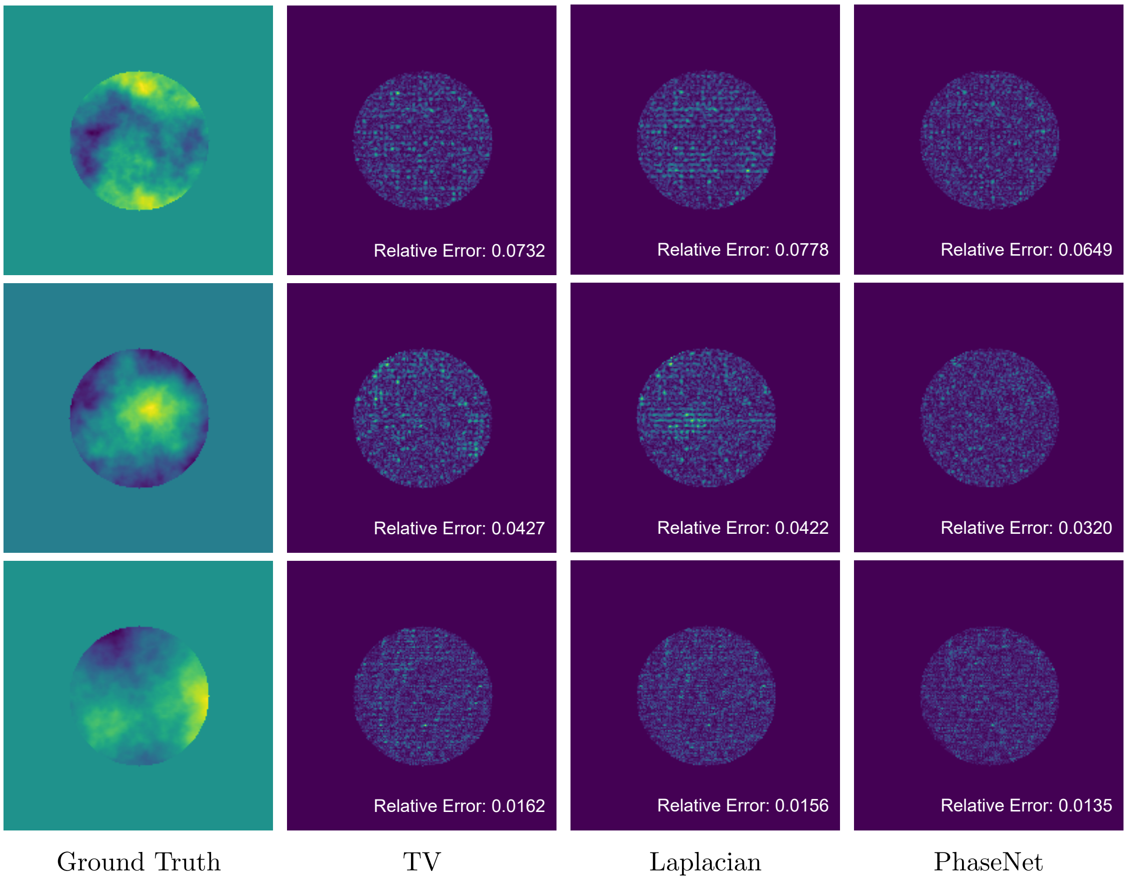


Figure 5: Visual comparison of phase reconstruction error images. The first row is from the Two-frame-three-layers dataset with seeing condition 40. The second row is from the Four-frame-three-layers dataset with seeing condition 40. The third row is from the Eight-frame-one-layer dataset with seeing condition 10.

where $g(\mathbf{x}, \mathbf{y})$ is the blurred observation, $k(\mathbf{x}, \mathbf{y})$ is the estimated PSF, and $f(\mathbf{x}, \mathbf{y})$ is the deblurred image. From the deconvolution results, we observe that the proposed method achieves the highest PSNR values compared to TV and Laplacian models and is closer to the true PSF results. Moreover, the deconvolution results obtained using PSFs reconstructed by TV and Laplacian methods introduce additional fluctuations in the restored image compared to our method, as seen in Figure 13.

4.4. Ablation study and discussion.

In this section, we evaluate the performance of our PhaseNet in the following perspectives.

Iteration number. We investigate the phase reconstruction performance with a different number of NAG steps in our PhaseNet on the Two-frames-one-layer dataset. The results are shown in Figure 7. From the table, we find the performance of 100 NAG steps is significantly worse than that of those models with more than 400 NAG steps. In addition, the relative

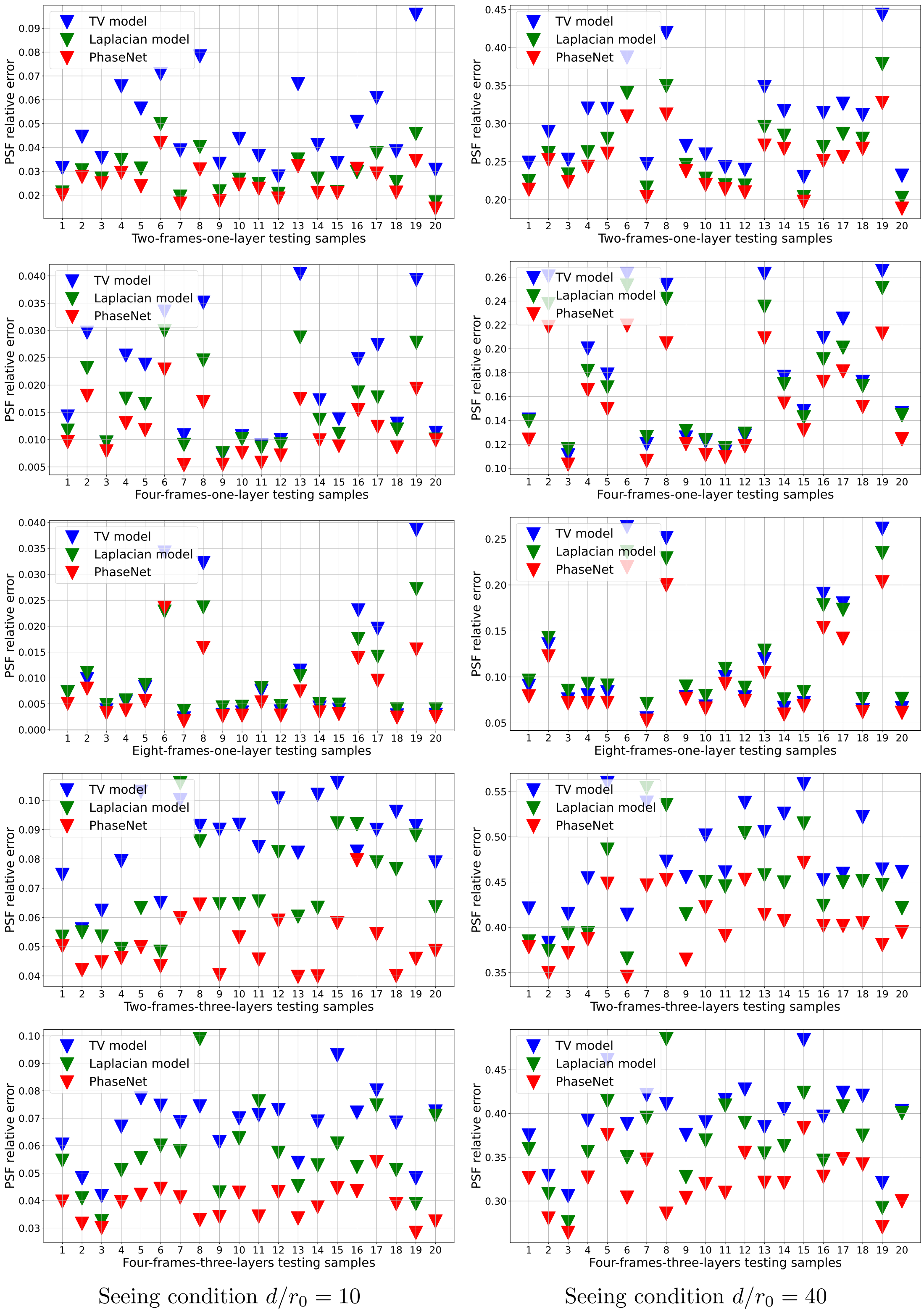


Figure 6: Comparison of relative errors of PSF on five testing datasets with two seeing conditions.

Table 6: Comparison of averaged relative errors of estimated PSF with different unrolling algorithms on the Two-frames-one-layer dataset.

Seeing condition	PhaseNet-GD	PhaseNet-ADMM	PhaseNet-NAG
10	0.0921	0.0290	0.0250
40	0.4078	0.2588	0.2459

Table 7: Comparison of averaged relative errors of estimated PSF with black box CNN solver on the Two-frames-one-layer dataset.

Seeing condition	RCAN	PhaseNet
10	0.2756	0.0250
40	0.7635	0.2459

Table 8: Comparison of averaged relative errors of estimated PSF on the Two-frames-one-layer dataset with different noise level.

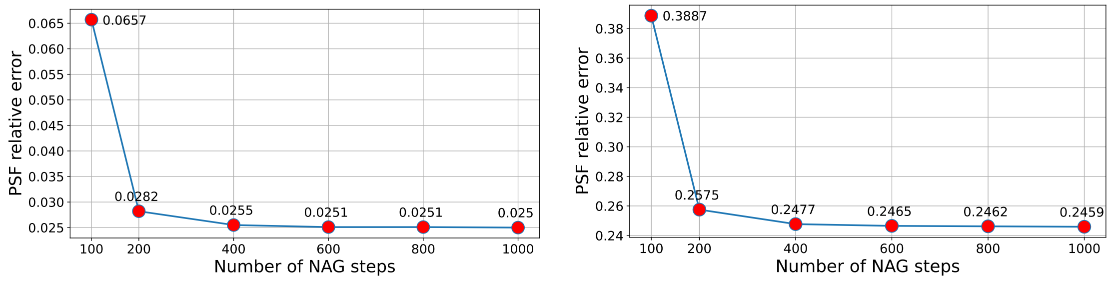
Seeing condition	Noise level	TV model	Laplacian model	PhaseNet (ours)
10	1 %	0.0489	0.0293	**0.0250**
	2 %	0.0492	0.0299	**0.0260**
	3 %	0.0495	0.0308	**0.0276**
40	1 %	0.3007	0.2638	**0.2479**
	2 %	0.3049	0.2677	**0.2541**
	3 %	0.3119	0.2738	**0.2640**

Table 9: Comparison of using all iteration output loss and final iteration output loss on the Two-frames-one-layer dataset.

Seeing condition	10	40
All iteration output loss	0.0258	0.2504
Final iteration output loss	0.0250	0.2459

Table 10: Comparison of averaged relative errors of estimated PSF on the Two-frames-one-layer dataset with different wind velocity relative error.

Seeing condition	RE in WV	TV model	Laplacian model	PhaseNet (ours)
10	10 %	0.0538	0.0402	**0.0374**
	20 %	0.0702	0.0571	**0.0494**
40	10 %	0.3612	0.3236	**0.3167**
	20 %	0.4413	0.4132	**0.3770**



Seeing condition $d/r_0 = 10$ Seeing condition $d/r_0 = 40$

Figure 7: Different number of NAG steps on the Two-frames-one-layer dataset.

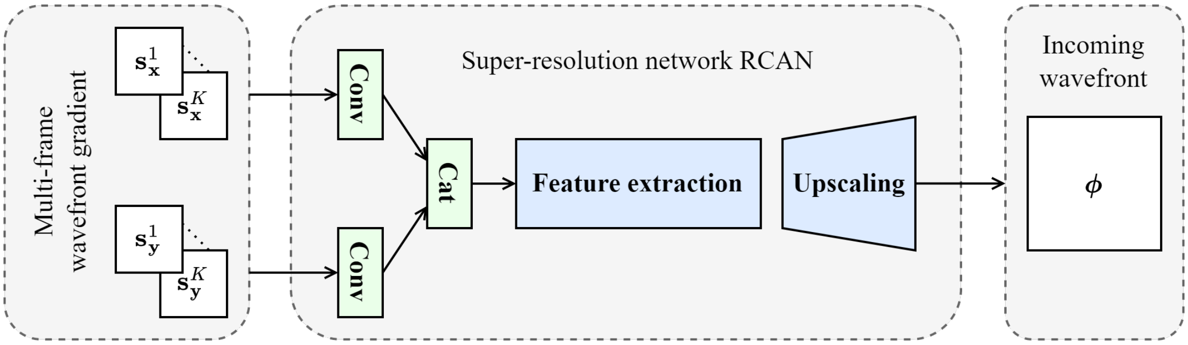
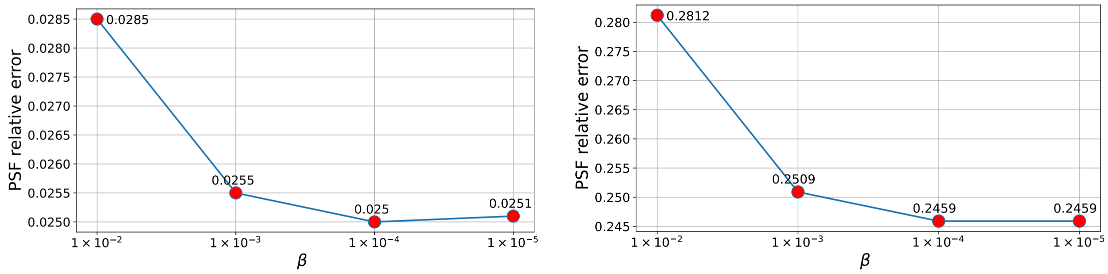


Figure 8: Phase reconstruction with an image super-resolution network RCAN.

error improves less on models with more than 800 NAG steps, and we use 1,000 NAG steps in our method.

Optimization algorithm. To validate the benefits of unrolling the NAG algorithm as our neural network, we compare the performance of unrolling the gradient descent algorithm and ADMM algorithm, which we refer to as PhaseNet-GD and PhaseNet-ADMM, with our PhaseNet-NAG method on the Two-frames-one-layer dataset. The iteration number of PhaseNet-GD is set to 1,000, which is consistent with PhaseNet-NAG. The implementation details for the PhaseNet-ADMM model are given in Appendix A. The results are shown in Table 6. From the table, we find the performance of PhaseNet-GD is significantly worse than PhaseNet-NAG and PhaseNet-ADMM. The result shows that the convergence rate of the GD algorithm is lower than the NAG algorithm and cannot be further improved through network training. Additionally, the results of PhaseNet-NAG are slightly better than PhaseNet-ADMM. One possible reason is that in one ADMM iteration, we use the CG algorithm to solve the first sub-problem, which will bring errors to the solution. In addition, PhaseNet-ADMM's computational overhead is also higher than PhaseNet-NAG, so we choose to unroll the NAG algorithm as our neural network.

Comparison with black box CNN solver. To demonstrate the advantages of using the unrolling method to design the neural network, we compare the phase reconstruction performance with an image super-resolution network, namely RCAN [51], which predicts the



Seeing condition $d/r_0 = 10$ Seeing condition $d/r_0 = 40$

Figure 9: Parameter analysis of β on Two-frames-one-layer dataset.

Table 11: Comparison of running time among different methods.

Data	TV model	Laplacian model	PhaseNet (ours)
Two-frames-one-layer	9.57s	1.82s	1.71s
Four-frames-one-layer	12.02s	3.22s	2.11s
Eight-frames-one-layer	14.32s	3.37s	2.24s
Two-frames-three-layers	82.79s	4.77s	10.56s
Four-frames-three-layers	133.85s	7.54s	17.21s

phase $\boldsymbol{\phi}$ from multi-frame wavefront gradient $\bar{\mathbf{s}}$ directly. In particular, we process the input wavefront gradients $\{\mathbf{s}_{\mathbf{x}}^i\}_{i=1}^K$ and $\{\mathbf{s}_{\mathbf{y}}^i\}_{i=1}^K$ through two convolution blocks, followed by merging the features of $\mathbf{s}_{\mathbf{x}}^i$ and $\mathbf{s}_{\mathbf{y}}^i$ via concatenation. Subsequently, the merged features are fed into the super-resolution network to recover the underlying incoming wavefront $\boldsymbol{\phi}$, as illustrated in Figure 8. We train the super-resolution network for 300,000 iterations with Adam [23] optimizer with batch size 8. The initial learning rate is set to 1×10^{-4} and is halved every 100,000 iterations. We use the PSF relative error as the loss function here to be consistent with PhaseNet. The results are shown in Table 7. The table shows that the super-resolution network's phase reconstruction performance is much worse than PhaseNet and traditional variational based methods. One possible reason is that the black box CNN solver does not preserve the mathematical structure in the original inverse problem, so the information from the forward model will be lost during the solving process. In addition, using unrolling can make our method more interpretable than a black box CNN solver.

Parameter analysis on β. We compare the phase reconstruction performance with different Laplacian regularization parameters. We train PhaseNet with different parameter β on the Two-frames-one-layer dataset, and the PSF relative error results are shown in Figure 9. From the results, we find the optimal setting for β is 1×10^{-4} or 1×10^{-5}, and we choose $\beta = 1\times10^{-4}$ in our PhaseNet models.

Restriction on the reconstruction process. One commonly used strategy in the deep unrolling method is to involve the output from all iterations in the final loss function:

$$\sum_{n=1}^{N} \ell(\boldsymbol{\phi}^n, \boldsymbol{\phi}_{true}) = \sum_{n=1}^{N} \frac{\|\mathbf{k}^n - \mathbf{k}_{true}\|_2}{\|\mathbf{k}_{true}\|_2}, \tag{4.7}$$

where $\boldsymbol{\phi}^n = \sum_{j=1}^{L} \boldsymbol{\phi}_j^n$, and $\mathbf{k}^n = |\mathcal{F}^{-1}\{\mathcal{P}\exp[\iota\boldsymbol{\phi}^n]\}|^2$. We compare the performance between using all iteration output loss (4.7) and the final stage output loss (3.14) on the Two-frames-one-layer dataset, see Table 9. The results show that the PSF error of using all iteration loss is slightly higher than that of only using final iteration loss. One possible reason is that the reconstructed phase at the first few steps is inaccurate. Involving all iteration outputs into the loss function will bring additional constraints to the reconstruction process, whereas using final iteration output loss leaves more freedom for the optimization process, which benefits the wavefront reconstruction.

Wavefront gradient noise level. We train our PhaseNet with different noise level wavefront gradient $\bar{\mathbf{s}}$ on the Two-frame-one-layer dataset. In particular, we assume the Gaussian noise level in the training dataset is 0% to 4%, and test with noise levels 1%, 2%, and 3%. The results are shown in Table 8. The results show that PhaseNet outperforms TV and Laplacian models on all noise levels and seeing conditions. Meanwhile, PhaseNet does not need to reselect regularization parameters for different noise levels, which is more convenient to use in practice than traditional methods.

Error in wind velocity. In our approach, the wind velocities are presumed to be known. As pointed out above, they may be captured with the assistance of balloons or additional instruments. To investigate the influence of a wrongly estimated wind velocity, we conducted an ablation study using the Two-frames-on-layer dataset. In our analysis, we trained PhaseNet with a wind velocity relative error range from 0% to 30%. We then tested our model on the wind velocity with relative errors of 10% and 20%. The results are displayed in Table 10. From the results, we deduced that PhaseNet consistently performed better than TV and Laplacian models across all noise levels and seeing conditions. This indicates that our method exhibits greater robustness against wind velocity errors.

Running time. The running time compared to the TV and Laplacian model is shown in Table 11. TV and Laplacian models are assessed on an Intel i5-10500 CPU. Our PhaseNet is examined on a single Nvidia GeForce RTX 3090 GPU. One of the main aspects influencing the time cost is the computation of the forward operator $\mathbf{H}$. Given the wavefront gradient frames and the wind velocity, calculating the $\mathbf{H}$ is necessitated, as seen in (3.1). This step is compute-intensive and thus impacts the overall runtime.

5. Conclusion.

In this work, we propose a deep learning based phase reconstruction model called PhaseNet for ground-based astronomy with multi-frame observations. The PhaseNet is constructed by unrolling the NAG algorithm to solve the traditional inverse problem. We adopt a neural network to approximate the residual between traditional Laplacian regularization and the unknown turbulence statistics. Compared with the traditional variational based method, the proposed PhaseNet achieves lower PSF relative errors among all atmosphere conditions.

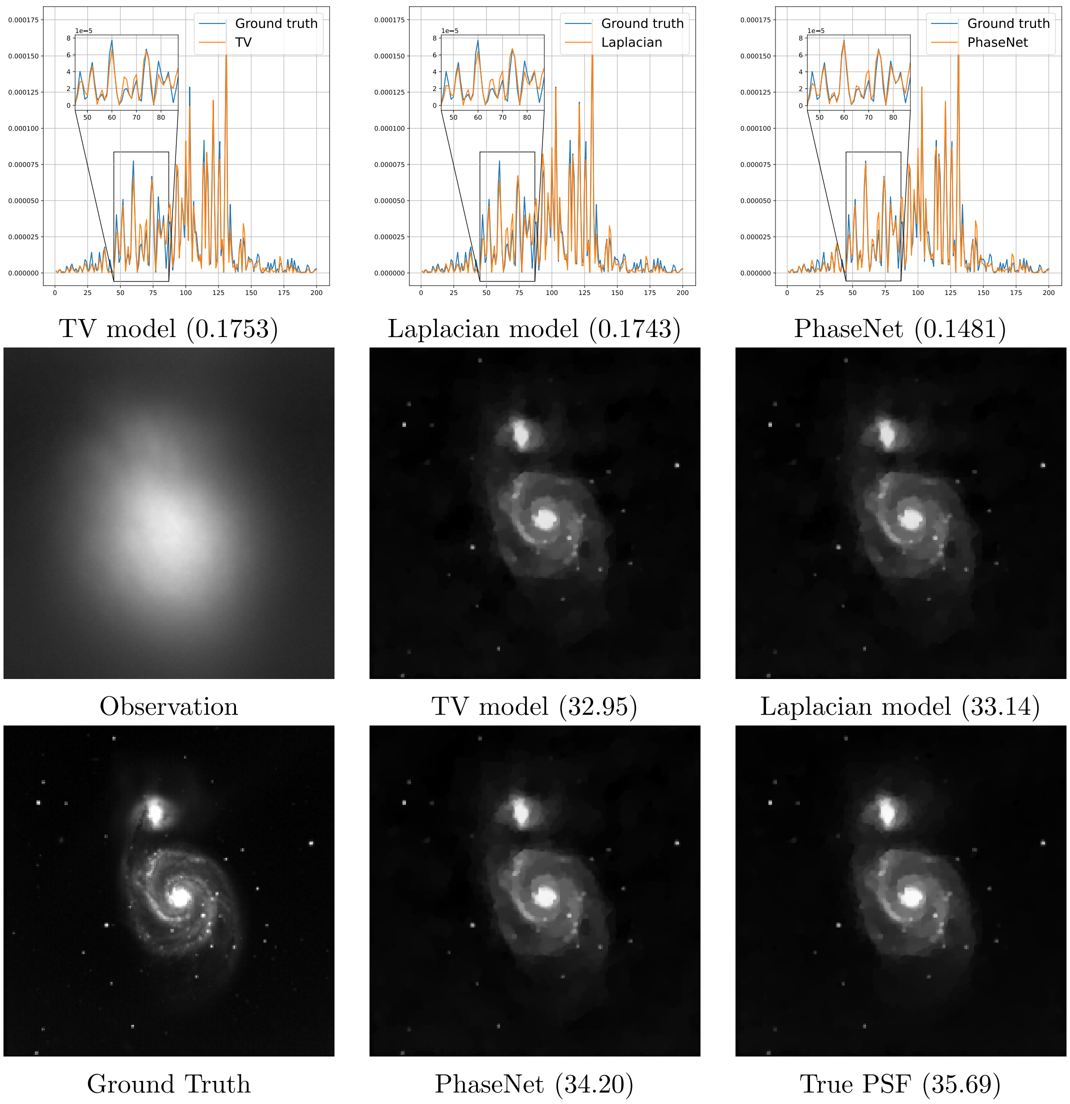
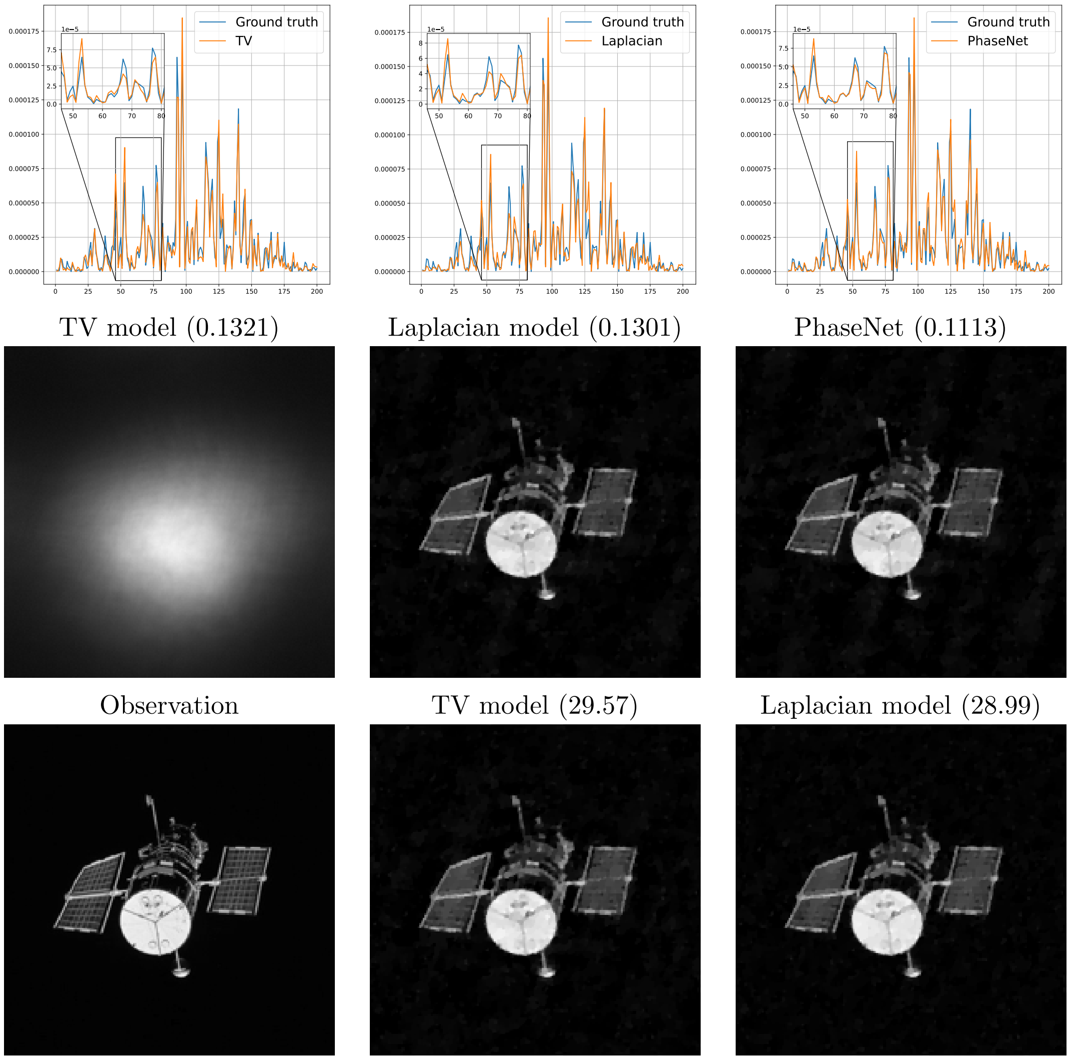


Figure 10: Visual comparison of estimated PSFs and deconvolution results on the Four-frames-one-layer dataset with seeing condition 40. The first row shows a cross-sectional comparison between the estimated and ground truth PSFs with PSF relative error. The second and third rows show the deconvolution results using different PSFs with PSNR.

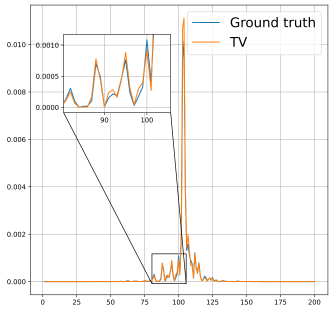


TV model (0.1321) | Laplacian model (0.1301) | PhaseNet (0.1113)

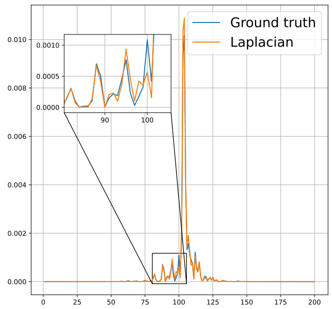
Observation | TV model (29.57) | Laplacian model (28.99)

Ground Truth | PhaseNet (30.47) | True PSF (31.30)

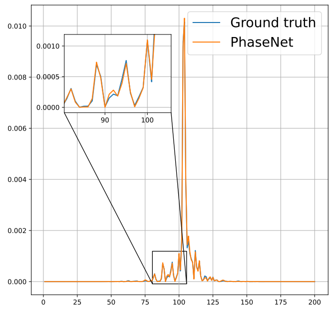
Figure 11: Visual comparison of estimated PSFs and deconvolution results on the Eight-frames-one-layer dataset with seeing condition 40. The first row shows a cross-sectional comparison between estimated and ground truth PSFs with PSF relative error. The second and third rows show the deconvolution results using different PSFs with PSNR.



TV model (0.1039)

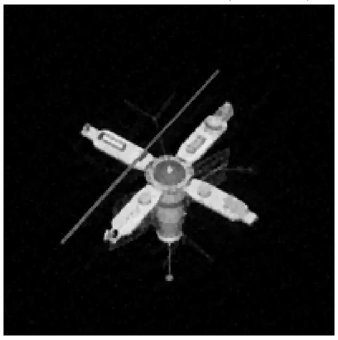


Laplacian model (0.1069)

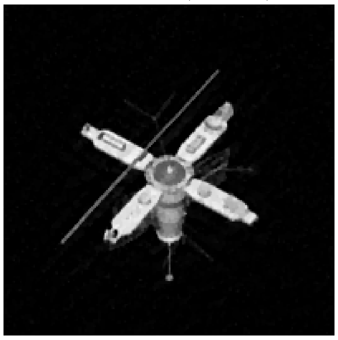


PhaseNet (0.0532)

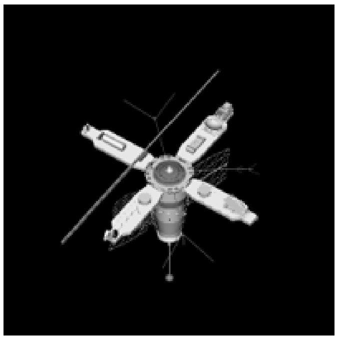
Observation



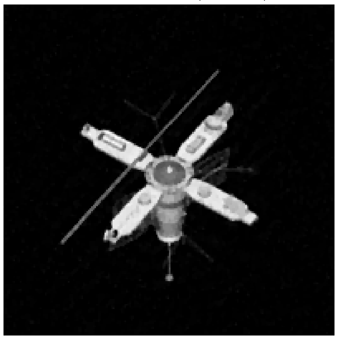
TV model (32.23)



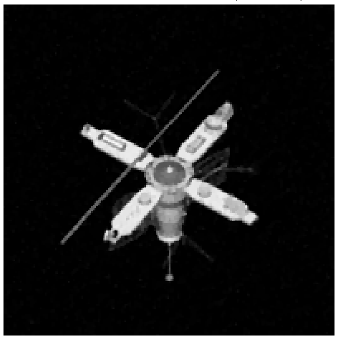
Laplacian model (32.68)



Ground Truth

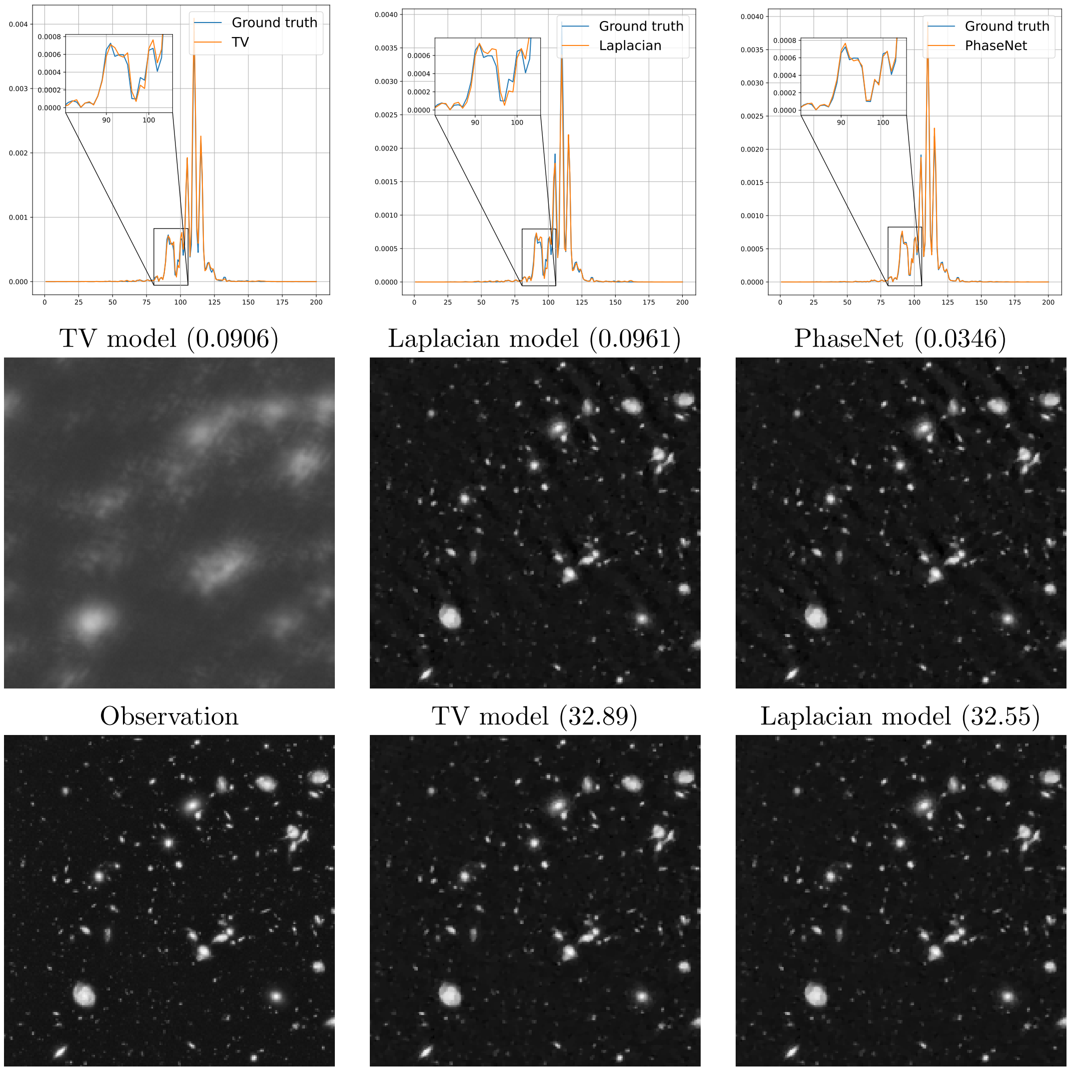


PhaseNet (34.49)



True PSF (35.10)

Figure 12: Visual comparison of estimated PSFs and deconvolution results on the Two-frames-three-layers dataset with seeing condition 10. The first row shows a cross-sectional comparison between estimated and ground truth PSFs with PSF relative error. The second and third rows show the deconvolution results using different PSFs with PSNR.



TV model (0.0906) | Laplacian model (0.0961) | PhaseNet (0.0346)

Observation | TV model (32.89) | Laplacian model (32.55)

Ground Truth | PhaseNet (34.49) | True PSF (34.68)

Figure 13: Visual comparison of estimated PSFs and deconvolution results on the Four-frames-three-layers dataset with seeing condition 10. The first row shows a cross-sectional comparison between estimated and ground truth PSFs with PSF relative error. The second and third rows show the deconvolution results using different PSFs with PSNR.

Algorithm .1 Forward propagation of PhaseNet-ADMM

Input: Multi-frame wavefront gradient $\bar{\mathbf{s}}$, $\mathbf{\Phi}^0$, ρ, number of iterations N.
Output: Reconstructed incoming wavefront $\boldsymbol{\phi} = \sum_{j=1}^{L} \boldsymbol{\phi}_j$.

1: Initialize $\hat{\mathbf{\Phi}}^0 = \mathbf{\Phi}^0$, $\mathbf{P} = 0$;
2: **for** $n = 0, 1, 2, \ldots, N$ **do**
3: Update $\mathbf{\Phi}^{n+1}$ in (A.4) with CG algorithm;
4: $\hat{\mathbf{\Phi}}^{n+1} = \mathbf{W}_\theta(\mathbf{\Phi}^{n+1} + \mathbf{P}^n/\rho)$;
5: $\mathbf{P}^{n+1} = \mathbf{P}^n + \rho(\mathbf{\Phi}^{n+1} - \hat{\mathbf{\Phi}}^{n+1})$;
6: **end for**
7: **return** Incoming wavefront $\boldsymbol{\phi} = \sum_{l=1}^{L} \boldsymbol{\phi}_l^{N+1}$.

Appendix A. Details on PhaseNet-ADMM model.

We provide more implementation details for the PhaseNet-ADMM model. Recall the objective function in our method is

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^{L}\gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2 + \mathbf{R}_\vartheta^{res}(\mathbf{\Phi}). \tag{A.1}$$

Introducing an auxiliary variable $\hat{\mathbf{\Phi}}$, we can reformulate the optimization problem in (A.1) as

$$\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^{L}\gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2 + \mathbf{R}_\vartheta^{res}(\hat{\mathbf{\Phi}}). \text{ s.t. } \hat{\mathbf{\Phi}} = \mathbf{\Phi}. \tag{A.2}$$

Then the augmented Lagrangian function for (A.2) is

$$\mathcal{L}(\mathbf{\Phi}, \hat{\mathbf{\Phi}}, \mathbf{P}) = \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^{L}\gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2 + \mathbf{R}_\vartheta^{res}(\hat{\mathbf{\Phi}}) + \frac{\rho}{2}\|\mathbf{\Phi} - \hat{\mathbf{\Phi}} + \mathbf{P}/\rho\|_2^2 - \frac{\rho}{2}\|\mathbf{P}/\rho\|_2^2, \tag{A.3}$$

where $\mathbf{P}$ is the dual variable and $\rho > 0$ is a chosen constant. The ADMM iteration is

$$\mathbf{\Phi}^{n+1} = \arg\min_{\mathbf{\Phi}} \frac{1}{2}\|\bar{\mathbf{s}} - \mathbf{H}\mathbf{\Phi}\|_2^2 + \frac{\beta}{2}\sum_{l=1}^{L}\gamma_l\|\mathbf{L}\boldsymbol{\phi}_l\|_2^2 + \frac{\rho}{2}\|\mathbf{\Phi} - \hat{\mathbf{\Phi}}^n + \mathbf{P}^n/\rho\|_2^2, \tag{A.4}$$

$$\hat{\mathbf{\Phi}}^{n+1} = \arg\min_{\hat{\mathbf{\Phi}}} \frac{\rho}{2}\|\mathbf{\Phi}^{n+1} - \hat{\mathbf{\Phi}} + \mathbf{P}^n/\rho\|_2^2 + \mathbf{R}_\vartheta^{res}(\hat{\mathbf{\Phi}}), \tag{A.5}$$

$$\mathbf{P}^{n+1} = \mathbf{P}^n + \rho(\mathbf{\Phi}^{n+1} - \hat{\mathbf{\Phi}}^{n+1}). \tag{A.6}$$

The first sub-problem (A.4) is a least-squares problem and can be solved through the CG algorithm. The second sub-problem (A.5) is equivalent to

$$\hat{\mathbf{\Phi}}^{n+1} = \mathrm{Prox}_{\rho\mathbf{R}_\vartheta^{res}}(\mathbf{\Phi}^{n+1} + \mathbf{P}^n/\rho), \tag{A.7}$$

where Prox denotes the proximal operator, which can be replaced by a neural network $\mathbf{W}_\theta$ and learn from the data. So (A.5) is equal to

$$\hat{\mathbf{\Phi}}^{n+1} = \mathbf{W}_\theta(\mathbf{\Phi}^{n+1} + \mathbf{P}^n/\rho). \tag{A.8}$$

We unroll the ADMM iteration as a neural network called PhaseNet-ADMM. In particular, we use 400 CG step to solve sub-problem (A.4), and the network architecture for network $\mathbf{W}_\theta$ is the same as $\mathbf{R}_\theta$ in our PhaseNet-NAG. We use 10 ADMM iterations in PhaseNet-ADMM and fix ρ as 10^{-4}.

REFERENCES

[1] C. Bao, H. Ji, Y. Quan, and Z. Shen, *Dictionary learning for sparse coding: Algorithms and convergence analysis*, IEEE transactions on pattern analysis and machine intelligence, 38 (2015), pp. 1356–1369.

[2] J. M. Bardsley, *Wavefront reconstruction methods for adaptive optics systems on ground-based telescopes*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 67–83.

[3] I. R. Bleyer and R. Ramlau, *A double regularization approach for inverse problems with noisy data and inexact operator*, Inverse Problems, 29 (2013), p. 025004.

[4] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.

[5] R. H. Chan, X. Yuan, and W. Zhang, *Point-spread function reconstruction in ground-based astronomy by l 1-l p model*, JOSA A, 29 (2012), pp. 2263–2271.

[6] R. H. Chan, X. Yuan, and W. Zhang, *A phase model for point spread function estimation in ground-based astronomy*, Science China Mathematics, 56 (2013), pp. 2701–2710.

[7] T. F. Chan and C.-K. Wong, *Total variation blind deconvolution*, IEEE transactions on Image Processing, 7 (1998), pp. 370–375.

[8] J. Choi, G. H. Soehnel, B. E. Bagwell, K. R. Dixon, and D. V. Wick, *Optical requirements with turbulence correction for long-range biometrics*, in Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI, vol. 7306, SPIE, 2009, pp. 382–392.

[9] Q. Chu, S. Jefferies, and J. G. Nagy, *Iterative wavefront reconstruction for astronomical imaging*, SIAM Journal on Scientific Computing, 35 (2013), pp. S84–S103.

[10] Q. Chu, S. Knepper, J. Nagy, and S. Jefferies, *Fast psf reconstruction using the frozen flow hypothesis*, in Signal Recovery and Synthesis, Optica Publishing Group, 2011, p. SMC4.

[11] T. B. DuBose, D. F. Gardner, and A. T. Watnik, *Intensity-enhanced deep network wavefront reconstruction in shack–hartmann sensors*, Optics Letters, 45 (2020), pp. 1699–1702.

[12] L. Dykes, R. Ramlau, L. Reichel, K. Soodhalter, and R. Wagner, *Lanczos-based fast blind deconvolution methods*, Journal of Computational and Applied Mathematics, 382 (2021), p. 113067, https://doi.org/10.1016/j.cam.2020.113067.

[13] B. L. Ellerbroek, *Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques*, JOSA A, 19 (2002), pp. 1803–1816.

[14] M. Eslitzbichler, C. Pechstein, and R. Ramlau, *An h1-kaczmarz reconstructor for atmospheric tomography*, Journal of Inverse and Ill-Posed Problems, 21 (2013), pp. 431–450.

[15] L. Gilles, *Order-n sparse minimum-variance open-loop reconstructor for extreme adaptive optics*, Optics Letters, 28 (2003), pp. 1927–1929.

[16] J. W. Goodman and P. Sutton, *Introduction to fourier optics*, Quantum and Semiclassical Optics-Journal of the European Optical Society Part B, 8 (1996), p. 1095.

[17] H. Guo, N. Korablinova, Q. Ren, and J. Bille, *Wavefront reconstruction with artificial neural networks*, Optics Express, 14 (2006), pp. 6456–6462.

[18] L. Hu, S. Hu, W. Gong, and K. Si, *Learning-based shack-hartmann wavefront sensor for high-order aberration detection*, Optics express, 27 (2019), pp. 33504–33517.

[19] L. Hu, S. Hu, W. Gong, and K. Si, *Deep learning assisted shack–hartmann wavefront sensor for direct wavefront detection*, Optics Letters, 45 (2020), pp. 3741–3744.

[20] S. M. Jefferies and M. Hart, *Deconvolution from wave front sensing using the frozen flow hypothesis*, Optics express, 19 (2011), pp. 1975–1984.

[21] L. Justen and R. Ramlau, *A general framework for soft-shrinkage with applications to blind deconvolution and wavelet denoising*, Applied and Computational Harmonic Analysis, 26 (2009), pp. 43–63.

[22] R. Ke, R. Wagner, R. Ramlau, and R. Chan, *Reconstruction of the high resolution phase in a closed loop adaptive optics system*, SIAM Journal on Imaging Sciences, 13 (2020), pp. 775–806.
[23] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980, (2014).
[24] A. N. Kolmogorov, *The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers*, Cr Acad. Sci. URSS, 30 (1941), pp. 301–305.
[25] A. N. Kolmogorov, *The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers*, Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, 434 (1991), pp. 9–13.
[26] E. Y. Lam and J. W. Goodman, *Iterative statistical approach to blind image deconvolution*, JOSA A, 17 (2000), pp. 1177–1184.
[27] Z. Li and X. Li, *Centroid computation for shack-hartmann wavefront sensor in extreme situations based on artificial neural networks*, Optics Express, 26 (2018), pp. 31675–31692.
[28] V. Monga, Y. Li, and Y. C. Eldar, *Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing*, IEEE Signal Processing Magazine, 38 (2021), pp. 18–44.
[29] J. Nagy, S. Jefferies, and Q. Chu, *Fast psf reconstruction using the frozen flow hypothesis*, in Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference, 2010, pp. 13–16.
[30] Y. Nesterov, *A method for solving the convex programming problem with convergence rate* $o(1/k^2)$, in Soviet Mathematics Doklady, 1983, pp. 372–376.
[31] A. Neubauer, *On the ill-posedness and convergence of the Shack-Hartmann based wavefront reconstruction*, J. Inv. Ill-Posed Problems, 18 (2010), pp. 551–576.
[32] A. Neubauer, *A new cumulative wavefront reconstructor for the shack–hartmann sensor*, Journal of Inverse and Ill-Posed Problems, 21 (2013), pp. 451–476.
[33] R. J. Noll, *Zernike polynomials and atmospheric turbulence*, JOsA, 66 (1976), pp. 207–211.
[34] B. C. Platt and R. Shack, *History and principles of shack-hartmann wavefront sensing*, 2001.
[35] L. A. Poyneer, B. A. Macintosh, and J.-P. Véran, *Fourier transform wavefront control with adaptive prediction of the atmosphere*, JOSA A, 24 (2007), pp. 2645–2660.
[36] R. Ramlau, K. M. Soodhalter, and V. Hutterer, *Subspace recycling–based regularization methods*, SIAM Journal on Matrix Analysis and Applications, 42 (2021), pp. 1480–1505.
[37] F. Roddier, *V the effects of atmospheric turbulence in optical astronomy*, in Progress in optics, vol. 19, Elsevier, 1981, pp. 281–376.
[38] M. C. Roggemann, B. M. Welsh, and B. R. Hunt, *Imaging through turbulence*, CRC press, 1996.
[39] O. Ronneberger, P. Fischer, and T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[40] J. Sun, H. Li, Z. Xu, et al., *Deep admm-net for compressive sensing mri*, Advances in neural information processing systems, 29 (2016).
[41] R. Swanson, M. Lamb, C. Correia, S. Sivanandam, and K. Kutulakos, *Wavefront reconstruction and prediction with convolutional neural networks*, in Adaptive Optics Systems VI, vol. 10703, SPIE, 2018, pp. 481–490.
[42] A. Talmi and E. N. Ribak, *Wavefront reconstruction from its gradients*, JOSA A, 23 (2006), pp. 288–297.
[43] G. I. Taylor, *The spectrum of turbulence*, Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, 164 (1938), pp. 476–490.
[44] E. Thiébaut and M. Tallon, *Fast minimum variance wavefront reconstruction for extremely large telescopes*, JOSA A, 27 (2010), pp. 1046–1059.
[45] T. von Kármán et al., *Mechanische Ähnlichkeit und Turbulenz*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1930 (1930), pp. 58–76.
[46] R. Wagner, O. Beltramo-Martin, C. Correia, R. Fétick, R. Ramlau, T. Fusco, and B. Neichel, *Overview of psf determination techniques for adaptive-optics assisted elt instruments*, in AO4ELT 6 Conference, 2019, http://ao4elt6.copl.ulaval.ca/proceedings/401-EE6b-241.pdf.
[47] H. Weichel, *Laser beam propagation in the atmosphere*, vol. 10319, SPIE press, 1990.
[48] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with*

applications to nonnegative tensor factorization and completion, SIAM Journal on imaging sciences, 6 (2013), pp. 1758–1789.

[49] M. Yudytskiy, T. Helin, and R. Ramlau, *Finite element-wavelet hybrid algorithm for atmospheric tomography*, JOSA A, 31 (2014), pp. 550–560.

[50] W. Zhang, *A phase model using the huber norm for estimating point spread function under frozen flow hypothesis*, Journal of Computational and Applied Mathematics, 397 (2021), p. 113657.

[51] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, *Image super-resolution using very deep residual channel attention networks*, in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286–301.

[52] M. Zhariy, A. Neubauer, M. Rosensteiner, and R. Ramlau, *Cumulative wavefront reconstructor for the shack-hartmann sensor*, in Signal Recovery and Synthesis, Optica Publishing Group, 2011, p. JTuA4.