# Res-U2Net: untrained deep learning for phase retrieval and image reconstruction

Carlos Osorio Quero,[1,*] Daniel Leykam,[2,4] and Irving Rondon Ojeda[3,5]

[1]*Computer Science Department, National Institute of Astrophysics, Optics and Electronics (INAOE), Puebla 72810, Mexico*
[2]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*
[3]*School of Computational Sciences, Korea Institute for Advanced Study (KIAS), Seoul 02455, Republic of Korea*
[4]*daniel.leykam@gmail.com*
[5]*irving.rondon@gmail.com*
*\*caoq@inaoep.mx*

**Conventional deep learning-based image reconstruction methods require a large amount of training data, which can be hard to obtain in practice. Untrained deep learning methods overcome this limitation by training a network to invert a physical model of the image formation process. Here we present a novel, to our knowledge, untrained Res-U2Net model for phase retrieval. We use the extracted phase information to determine changes in an object's surface and generate a mesh representation of its 3D structure. We compare the performance of Res-U2Net phase retrieval against UNet and U2Net using images from the GDXRAY dataset.** © 2024 Optica Publishing Group

## 1. INTRODUCTION

In recent times the field of computational imaging has witnessed significant advancements through the use of deep learning methods [1]. Deep learning has emerged as a promising approach for solving inverse problems encountered in computational imaging [2]. Groundbreaking studies have successfully demonstrated the effectiveness of deep learning for applications including optical tomography [3], 3D image reconstruction [4,5], phase retrieval [6,7], computational ghost imaging [8], digital holography [9–11], imaging through scattering media [12], fluorescence lifetime imaging under low-light conditions [13], unwrapping [14,15], and fringe analysis [16].

The deep learning-based artificial neural networks employed in computational imaging typically rely on a substantial collection of labeled data to optimize their weight and bias parameters through a training process [17]. This training enables the network to learn a universal function capable of mapping data from the object space to the image space. While traditional optimization methods struggle with this highly non-convex reconstruction problem, deep learning-based methods excel due to their nonlinear nature. Moreover, these methods can leverage statistical knowledge acquired from large datasets to infer solutions. Even though this reconstruction process is rapid in most cases [18], the training procedure can be time-consuming, lasting several hours or even days, depending on the network architecture and the volume of data employed. Furthermore, obtaining a large and diverse training dataset, which is essential for the effectiveness of neural networks, is challenging. This is particularly relevant to phase retrieval problems where

procuring an exhaustive collection of ground-truth images for training is frequently infeasible due to variability in the imaging apparatus. Limited training data then hampers the network's effectiveness and its capacity for generalization [19].

Recent advances in imaging applications have demonstrated the immense potential of unsupervised learning techniques, specifically those utilizing untrained networks [20–22]. By harnessing the inherent structure of neural networks without the need for training data, remarkable outcomes have been achieved. Two notable examples of untrained networks are the deep image prior [20] and deep decoder [23], which have effectively leveraged the network structure as a prior for image statistics, even without prior training. This approach involves employing a deep network with randomly initialized weights as an image generator to produce the recovered image. The network's weights are then repeatedly updated using a loss function that compares the generated image with the input data, such as a noisy image. This approach has demonstrated remarkable effectiveness in simulated image denoising [23], deblurring [24], phase retrieval [6,19], and super-resolution tasks [25].

However, in many computational imaging problems, the acquired measurements do not directly resemble the reconstructed image. Instead, a forward model governs the relationship between the scene and the measurements, incorporating the underlying physics of the image formation problem. In the case of phase retrieval, this model constructs the phase of a sample using a known intensity distribution pair at the object and measurement planes.

There are a variety of amplitude-based phase retrieval algorithms with varying complexity, generality, and resolution [26],

including the Fourier-Born and Gerchberg-Saxton methods [27,28]. The former provides a closed-form solution to the phase retrieval problem under the assumption that the object only scatters the light weakly, while the latter is a more general but iterative method. Iterative phase retrieval algorithms often face convergence issues because the inverse problem is ill-posed: the phase profile that reproduces a given set of amplitude measurements is not unique, and minor changes in the measured data may result in significant uncertainties in the estimated solution [29,30].

One way to mitigate these challenges is to replace iterative methods with methods based on gradient descent. While the phase retrieval problem is typically a non-convex optimization problem, with many sub-optimal local minima, neural network parameterizations of the solution such as those based on UNet or other convolutional neural networks can be trained to obtain good solutions [19,31]. The UNet architecture consists of an encoder-decoder structure, where the encoder captures the high-level features from the input intensity data, and the decoder reconstructs the phase from these features [31]. By incorporating physics-based constraints and priors into the network's design, a physics-enhanced deep neural network can learn to model the image formation process more accurately. This integration allows the network to exploit the known physics principles governing the imaging problem, thereby improving the rate of convergence compared to iterative phase retrieval methods such as the Gerchberg-Saxton algorithm and improving the quality and fidelity of the reconstructed images.

In this study, we consider a method for the reconstruction of 2D and 3D images by combining diffraction models with convolutional neural networks and mesh estimation of the phase image [32]. We consider three different neural networks for solving the inverse problem: UNet [31] and U2Net [33], which were previously used for unsupervised phase retrieval, as well as a new architecture called Res-U2Net. We show that the Res-U2Net-based architecture can achieve better performance compared to UNet and U2Net due to its ability to capture finer details of the image. Specifically, Res-U2Net incorporates additional layers in the form of downsampling and unsampling blocks, while preserving details with residual connections that skip some of the layers. To compare the performance of Res-U2Net against UNet and U2Net we use the GDXRAY [34] dataset of X-Ray images, which provides dimensional information. For the evaluation of the 2D image phase, we utilize standard metrics, namely no-reference image quality assessment (NR-IQA) [35], mean squared error (MSE) [36], and skewness. These metrics provide insights into the performance, processing time, and quality of the reconstructed 3D images [37].

The outline of this article is as follows: Section 2 briefly reviews prior works on phase retrieval, including the incorporation of deep learning-based techniques. Section 3 presents the architecture of our physics-informed Res-U2Net phase retrieval model, which we benchmark numerically in Section 4. We conclude with Section 5.

## 2. PHASE RETRIEVAL AND DEEP LEARNING

The goal of the phase retrieval problem is to reconstruct an object's near field profile $\psi$ estimate from intensity measurements $I$, given information about the characteristics of the imaging system, represented by a (possibly nonlinear) operator $A$ that represents the imaging system, relating the object to the measured intensities [38,39],

$$I = |A\psi|^2. \tag{1}$$

The absence of phase information makes the generic inverse problem challenging to solve [40], as there exist infinitely many possible solutions that can yield the same set of measured intensities [41],

$$\tilde{\psi} = \text{argmin}_\psi ||\tilde{I} - |A\psi|^2||_2, \tag{2}$$

where $\tilde{\psi}$ represents the estimated signal or image that we want to recover, $\psi$ represents the image whose phase information is to be retrieved, $\tilde{I}$ is the measured intensity, and $\text{argmin}_\psi$ indicates that we are looking for the argument (in this case, $\tilde{\psi}$) that minimizes the following expression. Estimation of the signal $\tilde{\psi}$ given $I$ can be posed as a non-convex optimization problem [42].

In many imaging setups, $A$ will simply be a matrix encoding the two-dimensional Fourier transform, which forms the basis for the Fourier-Born and Gerchberg-Saxton methods, among others. In this case, $I = I_z(x, y)$ is the intensity in the imaging plane $(x, y)$ at a distance $z$ from the object, which is related to the near field $\psi_0(x, y)$ via the transfer function $H_z$,

$$I_z(x, y) = \left| \mathcal{F}^{-1}[e^{-ikz\sqrt{1-\lambda^2(k_x^2+k_y^2)}}\mathcal{F}[\psi_0(x, y)]] \right|^2 = |H_z\psi_0(x, y)|^2, \tag{3}$$

where $k = 2\pi/\lambda$, $\lambda$ is the imaging wavelength, and $\mathcal{F}$ is the two-dimensional Fourier transform. In the case of uniform illumination of a pure phase object, the near field can be written as $\psi_0(x, y) = I_0 e^{i\theta(x,y)}$, and the aim is to determine the phase profile $\theta(x, y)$ using $I_z(x, y)$.

In the case of iterative phase retrieval methods, such as the Gerchberg-Saxton algorithm, one iteratively updates the estimated image $\tilde{\psi} \in \mathbb{C}^{N \times N}$ in the spatial and Fourier domains [27]. However, these kinds of iteration procedures have several drawbacks. They tend to stagnate and exhibit slow convergence, often requiring more than 1000 iterations to reach a solution. In addition, they are highly sensitive to the initial conditions.

Alternatively, if the object is known to scatter the probe light only weakly (Born approximation), the near field can be written as $\psi_0(x, y) \approx I_0(1 + i\theta(x, y))$, and it follows that

$$I_z(x, y) = I_0(1 + 2\text{Re}[H_z\theta(x, y)]), \tag{4}$$

allowing the near field to be directly reconstructed from the Fourier transform of the far-field intensity. For more information about Fourier phase retrieval algorithms, we recommend interested readers to consult Refs. [3,27,28].

To address the limitations of conventional phase retrieval methods, newer gradient-based algorithms such as Writinger flow (WF) phase retrieval have been introduced [43], which solve the phase retrieval problem Eq. (1) using gradient descent [44],

$$\tilde{\psi}^{j+1} = \tilde{\psi}^j - \mu^{j+1}\nabla f(\tilde{\psi}^j), \tag{5}$$

where $\nabla f(\tilde{\psi}^j)$ represents the first-order gradient of the loss function and $\mu^{j+1}$ denotes the step size at the current iteration $j$. WF offers theoretical guarantees for convergence to the globally minimal solution. Empirically, a set of 4 to 8 intensity iterations is usually enough for convergence to the globally optimal solution. Unfortunately, WF often fails to converge to a satisfactory outcome when provided with only a single intensity measurement, limiting its effectiveness in this case unless one can impose prior constraints on the form of the object.

Recently extensive research has been conducted on phase retrieval using deep learning-based approaches. These approaches offer faster, non-iterative inferences compared to other more time-consuming optimization-based algorithms [26]. Most of these methods can reconstruct the phase using a single Fourier intensity measurement without any additional constraints. This seemingly ill-posed problem can have a unique solution (with only minor ambiguities) for the original complex-valued signal if the Fourier intensity measurement is oversampled by at least twice in each dimension [26,45]. In general, deep learning-based unsupervised phase retrieval methods can be categorized based on whether or not they incorporate the underlying physics into their networks [46].

The first category of deep learning-based methods involves using a feedforward network to directly estimate target images from a Fourier intensity measurement [47]. For instance, Ref. [48] proposed a two-branch convolutional neural network (CNN) to reconstruct the magnitude and phase of the 3D crystal image from an oversampled Fourier intensity measurement. While this approach can demonstrate reasonably good performance for simple images with limited details, its efficacy for complex images remains unknown. Other image reconstruction methods based on conditional generative adversarial networks [49] and multiple multilayer perceptrons [50] tailored for Fourier phase retrieval tasks can similarly capture simple features adequately but give relatively large errors in capturing finer details [51].

The second category of deep learning-based phase retrieval methods aims to enhance the quality of reconstructed images by effectively leveraging the underlying physics into their models [52]. One such approach incorporates physics information through a learnable spectral initialization [53] followed by a double-branch UNet for reconstruction. However, this approach requires an additional masking scheme to impose constraints on the measurement, leading to rather noisy reconstructed images, even for simple images [54]. Another proposed method involves the use of multi-layer perceptrons of different sizes in a cascaded network. In this approach, the intensity measurement is applied to each multi-layer perceptron to assist in training and inference [55]. Despite its merits, this approach struggles to reconstruct fine image details and necessitates a large network size due to the utilization of multiple multi-layer perceptrons [56].

To enhance the performance of the neural network model, it is beneficial to integrate physics-informed principles into a deep learning framework. This involves training the neural network to effectively learn the inverse mapping from the observed intensity data $I_z(x, y)$ to the corresponding near-field phase profile $\theta(x, y)$ using a forward diffraction model [31,57–59].

The estimated phase profile is then used as input into the diffraction model to obtain an estimate for the far-field intensity; the difference between this estimate and the measured intensity serves as a loss function used to train the network and improve the phase estimate. UNet, a popular architecture in the field of image processing and computer vision, can learn the inverse of the diffraction operator [19,57,60,61].

By integrating a UNet architecture within the phase retrieval process, we can harness the powerful representational capabilities of deep neural networks while retaining the convergence characteristics typical of traditional gradient descent-based phase retrieval algorithms. This UNet-based approach excels in capturing the complex interactions between intensity and phase data, rooted in the underlying physics of the problem. Additionally, the process involves a progressive refinement aspect, distinct from iterative methods in numerical analysis. This refinement, occurring over multiple cycles, allows for gradual improvement of the phase estimate. Such an approach is particularly beneficial in mitigating the challenges posed by initial conditions and noise in the intensity data, thereby enhancing the accuracy of phase retrieval [60].

## 3. PHASE RETRIEVAL USING RES-U2NET

The untrained phase retrieval process (see Fig. 1) involves applying a Fourier-based forward model [26], which is defined in Section 2, to evaluate the input image and obtain the image plane intensity $I_z(x, y)$. This diffraction model is then used as input for an untrained neural network that estimates the near-field phase $\tilde{\theta}(x, y)$. We evaluate three different neural networks: UNet, U2Net, and Res-U2Net. The neural network is trained by comparing the intensity profile $\tilde{I}_z(x, y)$ obtained from the diffraction model $A$, specified by Eq. (3) or Eq. (4), using the estimated phase $\tilde{\theta}(x, y)$, and comparing against the measured far-field intensity $I_z(x, y)$. The neural network is trained by minimizing the mean square error (MSE) between the measured and estimated intensities, $||I_z(x, y) - \tilde{I}_z(x, y)||$, following the approach introduced in Ref. [6]. This minimization of the cost function via gradient descent allows for the gradual refinement of the estimated phase until a desired accuracy, as measured by the difference between subsequent iterations, is obtained.
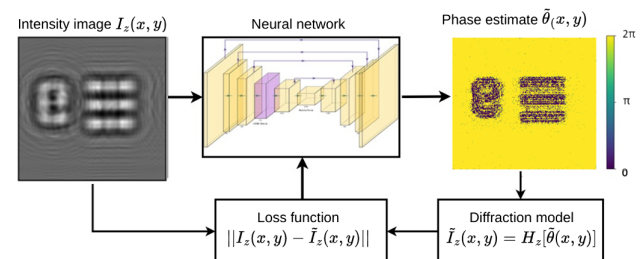


**Fig. 1.** Schematic of the phase retrieval process. An intensity image $I_z(x, y)$ is fed into a neural network, returning an estimate of the near-field phase $\tilde{\theta}(x, y)$. Diffraction model $H_z$ converts the estimated near-field phase to an estimated far-field intensity profile $\tilde{I}_z(x, y)$. The mean square error (MSE) between $I_z(x, y)$ and $\tilde{I}_z(x, y)$ serves as a loss function for optimizing the parameters of the neural network.
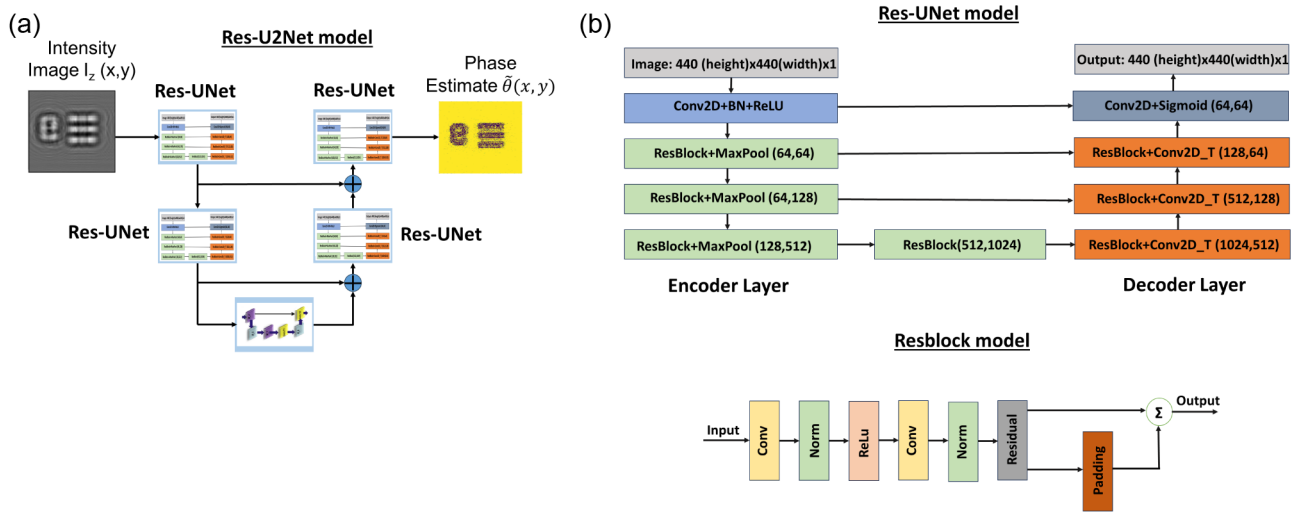
**Fig. 2.** Res-U2Net architecture: (a) U2Net model configuration, based on a multi-scale sequence of Res-UNet models, (b) Res-UNet model, the encoder extracts features using convolutional layers (Conv2D) with batch normalization, ReLU activation (ResBlock), and spatial resolution reduction via max pooling (MaxPooling2D). This is followed by a decoder assigning phases to the features by upsampling using transpose convolutions (Conv2DTranspose) with skip connections. Residual connections link the encoder and decoder layers to improve the training performance. Finally, a $1 \times 440 \times 440$ convolutional layer generates the segmentation mask, resulting in the network output.

## A. Res-U2Net Structure

The Res-U2Net architecture, as shown in Fig. 2, is a sophisticated evolution of the UNet and Res-UNet models, particularly designed for image segmentation tasks [31]. This architecture innovatively incorporates residual connections at various stages, enhancing information exchange and gradient flow during training. Structurally similar to UNet, it includes both downsampling and upsampling processes. The downsampling path consists of convolutional layers with batch normalization and ReLU activation, followed by max-pooling layers. Conversely, the upsampling path employs transposed convolutional layers to enlarge the feature maps. A key advancement of Res-U2Net over UNet is the amalgamation of a series of encoder/decoder modules that combine the Res-Unet model (refer to Fig. 2) with a series of stacked U-Nets. This configuration, featuring layers for both downsampling and upsampling, promotes more efficient feature transmission and mitigates the issue of vanishing gradients [62].

Res-U2Net has shown exceptional capabilities in various image segmentation tasks, surpassing the original UNet's performance. While its design is primarily for image segmentation, this does not directly imply its effectiveness in phase retrieval tasks, since phase retrieval often involves complex patterns rather than areas of near-uniform phase values, challenging the assumption that segmentation-focused methods would automatically excel here. Nevertheless, the U-Net architecture has been successfully applied to the phase retrieval [6]. However, we aim to explore new possible architectures using familiar ones.

Res-U2Net performs the following standard sequence of operations [see Fig. 2(a)]:

1. Res-UNet: The architecture applies a series of Res-UNet blocks in parallel, allowing the network to process and extract features at different scales or levels of abstraction.

2. Combination: The features extracted by the parallel Res-UNet blocks are combined to produce a richer set of features.

In Fig. 2(b), it is shown how Res-UNet carries out the encoder and decoder operations:

1. Feature Extraction (Encoder layer): This part of the model uses convolutional layers, batch normalization, and ReLU activation functions to process the input image. The convolutional layers are designed to extract features by applying filters that capture spatial hierarchies in the image. As the input passes through these layers, it is transformed into a set of feature maps that represent different aspects of the input. The batch normalization helps in stabilizing the learning process by normalizing the input of each layer, and the rectified linear unit (ReLU) activation function introduces non-linearity, allowing the model to learn complex patterns.

2. Spatial Resolution Reduction: Max pooling is used after convolutional blocks to reduce the spatial dimensions of the feature maps. This operation helps in making the representation smaller and more manageable, and it also introduces some level of invariance to small translations in the input image.

3. Upsampling (Decoder Layer): The decoder part of the network uses transpose convolutions (also known as up-convolutions or deconvolutions) to increase the spatial dimensions of the feature maps. This process is essential for tasks like image segmentation, where the goal is to produce an output image that is the same size as the input image.

4. Skip Connections: These connections are used to combine feature maps from the downsampling path with those from the upsampling path. By doing so, the network can leverage both high-level semantic information and low-level spatial

information, which is crucial for accurately reconstructing the output.

5. Residual Connections: After concatenating feature maps from different layers, additional convolutions are applied, and their output is added element-wise to the input of the concatenation. This creates a residual block (Resblock), which helps mitigate the vanishing gradient problem and allows for deeper networks by promoting more effective backpropagation of gradients.

6. Upsampling and Concatenation: This step repeats the process of upsampling and combining feature maps from different layers of the network. It ensures that as the spatial dimensions are restored to match the input size, the network progressively refines the details of the output.

7. Segmentation Mask Generation: Finally, we introduced a convolutional layer with a sigmoid activation function to generate the final segmentation mask, serving as the output of the neural network. This step converts the input vector into a matrix, containing the estimated phase information. The resulting matrix matches the dimensions of the input image, specifically a size of $1 \times 440 \times 440$.

### B. 3D Phase Reconstruction

We utilized the unified shape-from-shading model (USFSM) to perform the 3D reconstruction of the estimated image obtained through phase retrieval. The USFSM approach constructs three-dimensional representations by analyzing the spatial intensity variations present in the two-dimensional recovered image [63]. To extract depth information from the phase retrieval image, which corresponds to the surface points of the scene, we employed the fast-sweeping method. This method employs the Lax–Friedrichs Hamiltonian technique [64] to solve for the surface, utilizing an iterative sweeping strategy based on the fast sweeping scheme described in Ref. [32]. An example is shown in Fig. 3.

## 4. RESULTS

We carried out numerical calculations to compare the performance of phase retrieval using Res-U2Net against the UNet and U2Net networks, considering Fourier and Fourier-Born diffraction models and four $440 \times 440$ pixel images from the GDXRAY dataset, shown in Fig. 4. In all cases, we limit the maximum number of training iterations to 1000. Training was performed using the Keras framework in Python [65], and a stopping criterion with a network error tolerance of $10^{-4}$. The calculations were executed using an NVIDIA GTX 1080 graphics processing unit (GPU).
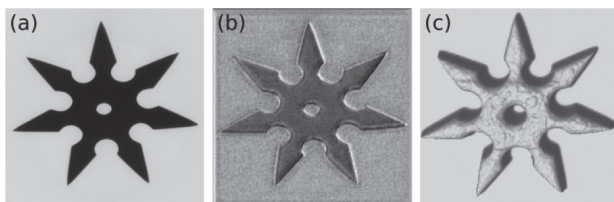


**Fig. 3.** 3D phase retrieval: (a) 2D Ray-X test image, (b) 2D phase retrieval estimate, and (c) resulting 3D mesh.
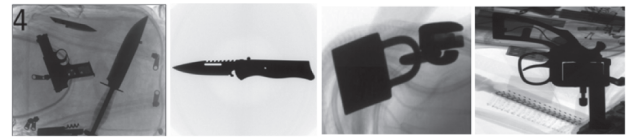


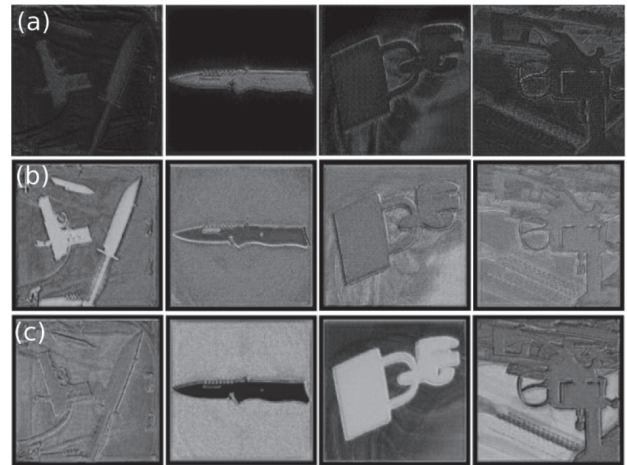**Fig. 4.** Examples from the GDXRAY dataset of $440 \times 440$ pixel images.



**Fig. 5.** 2D Fourier phase retrieval using (a) UNet, (b) U2Net, and (c) Res-U2Net.

The performance of the trained models was evaluated based on the quality of the reconstructed 2D images and 3D meshes. The configurations of the neural network and diffraction models were analyzed to determine the most efficient approach in terms of image, processing time, and mesh reconstruction quality.

### A. 2D Phase Retrieval

Figures 5 and 6 present the estimated phase profiles obtained from the different neural networks using the Fourier and Fourier-Born diffraction models, respectively. The image processing times ranged from 0.5 to 5 s for images of differing
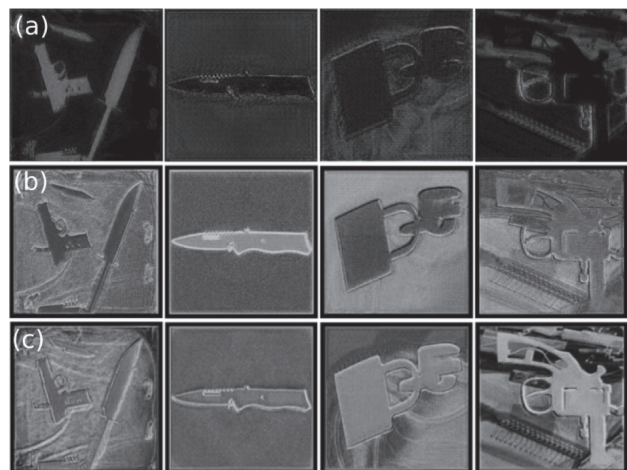


**Fig. 6.** 2D Fourier-Born phase retrieval using (a) UNet, (b) U2Net, and (c) Res-U2Net.

**Table 1.    Performance of 2D Phase Retrieval Using UNet, U2Net, and Res-U2Net Quantified Using Blind/Referenceless Image Spatial Quality Evaluator (Lower Is Better) [66]**

| Method | UNet | U2Net | Res-U2Net |
|---|---|---|---|
| Fourier | 16.37 | 11.67 | 9.63 |
| Fourier-Born | 15.17 | 9.04 | 8.12 |

**Table 2.    Performance of 2D Phase Retrieval Using UNet, U2Net, and Res-U2Net Quantified Using Natural Image Quality Evaluator (Lower Is Better) [67]**

| Method | UNet | U2Net | Res-U2Net |
|---|---|---|---|
| Fourier | 3.23 | 2.89 | 1.69 |
| Fourier-Born | 2.62 | 2.53 | 1.41 |

complexity. The networks differ in the detail and contrast of the phase images produced.

To quantify the performance of the phase reconstruction we used NR-IAQ, considering first the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) method [66]. This method assesses the statistical properties of an image to estimate its quality, with lower scores indicating better image quality and higher scores indicating poorer quality. The BRISQUE scores presented in Table 1 show that Res-U2Net consistently outperforms UNet and U2Net for both the generic Fourier imaging and the special case of weak phase contrast (Fourier-Born approximation). Res-U2Net consistently achieves the lowest BRISQUE scores, indicative of its superior ability to produce images with the highest perceived quality.

Next, we consider the Natural Image Quality Evaluator (NIQE) [67], which assesses factors such as texture, sharpness, and entropy. Higher values of NIQE indicate images with lower perceptual quality, while lower values suggest higher quality. A similar trend to BRISQUE is visible in the NIQE scores presented in Table 2. Once again, Res-U2Net achieves the lowest scores across all image types, signifying superior image quality and enhanced contrast levels, as can be seen in Figs. 5 and 6.

### B.  3D Retrieval Phase

Next we applied the shape-from-shading model (USFSM) to reconstruct 3D images from the 2D phase profiles obtained from the Fourier and Fourier-Born diffraction models. The resulting images are presented in Figs. 7 and 8.

To quantify the performance of the 3D image reconstruction, we calculated the mean squared error (MSE) and skewness [36], which defines the symmetry of the 3D shapes. A skewness value near 0 indicates the best mesh, and a value close to 1 indicates a completely degenerate mesh [68] between the 3D mesh of the normalized test images (Fig. 4) and the 3D mesh obtained from the phase reconstruction. The results can be found in Tables 3 and 4.

Inspecting the MSE values, UNet has a somewhat worse performance compared to the other two neural networks for both diffraction models. Moving on to the skewness values, again UNet has the worst performance, indicating lower spatial resolution in the reconstructed 3D image, limiting its ability to
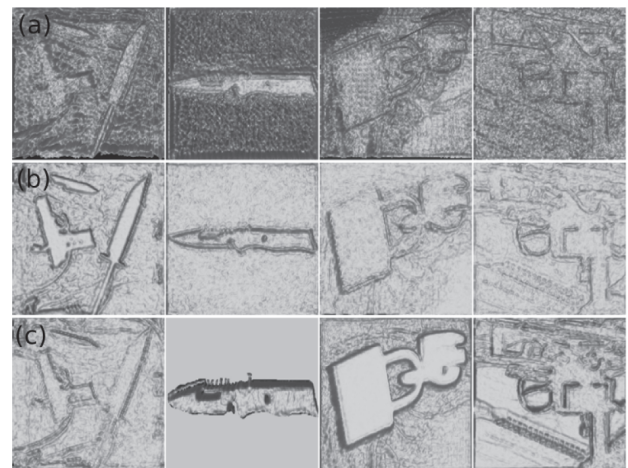


**Fig. 7.**    3D Fourier phase retrieval images obtained using (a) UNet, (b) U2Net, and (c) Res-U2Net.
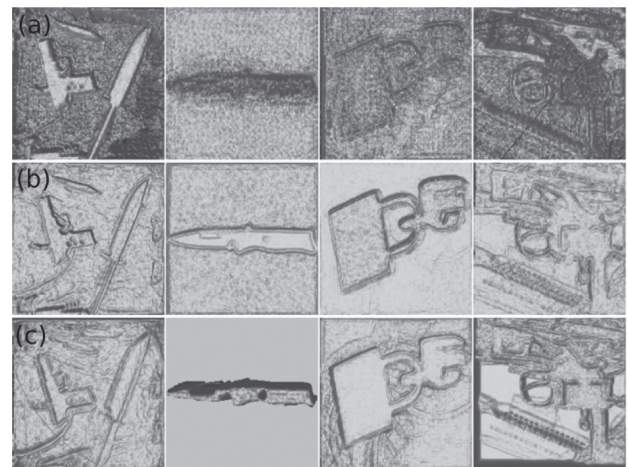


**Fig. 8.**    3D Fourier-Born phase retrieval images using (a) UNet, (b) U2Net, and (c) Res-U2Net.

**Table 3.    Performance of 3D Image Reconstruction from the UNet, U2Net, and Res-U2Net Phase Images Quantified by the Mean Squared Error (Lower Is Better)**

| Method | UNet | U2Net | Res-U2Net |
|---|---|---|---|
| Fourier | 0.150 | 0.059 | 0.056 |
| Fourier-Born | 0.270 | 0.062 | 0.053 |

**Table 4.    Performance of 3D Image Reconstruction from the UNet, U2Net, and Res-U2Net Phase Images Quantified by the Skewness (Lower Is Better)**

| Method | UNet | U2Net | Res-U2Net |
|---|---|---|---|
| Fourier | 0.545 | 0.1460 | 0.1140 |
| Fourier-Born | 1.145 | 0.0100 | 0.0049 |

discern surface details of the tested objects, as is visible in Figs. 7 and 8. Meanwhile, U2Net and Res-U2Net exhibit significantly better skewness values for the Fourier-Born diffraction model. We also note that methods incorporating physics-based forward

diffraction models tend to yield lower skewness values compared to other neural network models.

## 5. CONCLUSION

In this study, we examined the use of physics-informed deep learning techniques for phase retrieval, considering the specific example of phase retrieval from X-ray images. Our main objective was to assess the efficacy of these methods in 2D and 3D imaging. We conducted a thorough analysis of three neural networks—UNet, U2Net, and Res-U2Net—to determine their suitability for unsupervised Fourier phase retrieval in X-ray imaging. Our findings reveal significant improvements in both 2D and 3D reconstructions, with processing times ranging from 0.5 to 5 s. Notably, we observed enhanced background details and improved skewness scores in 3D meshes generated from GDXRAY test images. Res-U2Net, in particular, exhibited promising potential as a robust method for producing high-quality 2D Fourier phase retrieval images. On the other hand, UNet may require further optimizations to match the effectiveness of its counterparts. This investigation highlights the synergy between neural network models and physics-based forward models, offering an effective approach for phase retrieval tasks in 3D mesh normalized test images.

Future research should focus on exploring additional evaluation metrics and refining these models to optimize their performance in specific imaging applications. For example, the integration of generative adversarial networks (GANs) may be used to increase the network's robustness against noise and decrease the number of artifacts incorporated during the image phase estimation [49]. Res-U2Net can be adapted for phase retrieval of images in other spectral bands, and it can also be applied to other imaging problems found in different fields, such as biomedical imaging.

**Disclosures.** The authors declare no conflicts of interest.

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

## REFERENCES

1. D. Zhang and Z. Tan, "A review of optical neural networks," Appl. Sci. **12**, 5338 (2022).
2. S. Liu, X. Chen, T. Yang, *et al.*, "Machine learning aided solution to the inverse problem in optical scatterometry," Measurement **191**, 110811 (2022).
3. A. Burvall, U. Lundström, P. A. C. Takman, *et al.*, "Phase retrieval in X-ray phase-contrast imaging suitable for tomography," Opt. Express **19**, 10359–10376 (2011).
4. X.-F. Han, H. Laga, and M. Bennamoun, "Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era," IEEE Trans. Pattern Anal. Mach. Intell. **43**, 1578–1604 (2021).
5. L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an LED array microscope," Optica **2**, 104–111 (2015).
6. F. Wang, Y. Bian, H. Wang, *et al.*, "Phase imaging with an untrained neural network," Light Sci. Appl. **9**, 77 (2020).
7. Y. Shechtman, Y. C. Eldar, O. Cohen, *et al.*, "Phase retrieval with application to optical imaging: a contemporary overview," IEEE Signal Process. Mag. **32**(3), 87–109 (2015).
8. T. Shimobaba, Y. Endo, T. Nishitsuji, *et al.*, "Computational ghost imaging using deep learning," Opt. Commun. **413**, 147–151 (2018).
9. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: a review," Opt. Express **29**, 40572–40593 (2021).
10. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," Opt. Express **26**, 22603–22614 (2018).
11. Z. Ren, Z. Xu, and E. Y. M. Lam, "End-to-end deep learning framework for digital holographic reconstruction," Adv. Photonics **1**, 016004 (2019).
12. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**, 13738–13743 (2016).
13. J. T. Smith, N. Un, R. Yao, *et al.*, "Fluorescent lifetime imaging improved via deep learning," in *Biophotonics Congress: Optics in the Life Sciences Congress* (Optica Publishing Group, 2019), paper NM3C.4.
14. K. Wang, Y. Li, Q. Kemao, *et al.*, "One-step robust deep learning phase unwrapping," Opt. Express **27**, 15100–15115 (2019).
15. K. Wang, J. Di, Y. Li, *et al.*, "Transport of intensity equation from a single intensity image via deep learning," Opt. Laser Eng. **134**, 106233 (2020).
16. S. Feng, Q. Chen, G. Gu, *et al.*, "Fringe pattern analysis using deep learning," Adv. Photonics **1**, 025001 (2019).
17. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921–943 (2019).
18. R. Shang, K. Hoffer-Hawlik, F. Wang, *et al.*, "Two-step training deep learning framework for computational imaging without physics priors," Opt. Express **29**, 15239–15254 (2021).
19. K. Monakhova, V. Tran, G. Kuo, *et al.*, "Untrained networks for compressive lensless photography," Opt. Express **29**, 20913–20929 (2021).
20. V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 9446–9454.
21. K. Wang, J. Dou, Q. Kemao, *et al.*, "Y-net: a one-to-two deep learning framework for digital holographic reconstruction," Opt. Lett. **44**, 4765–4768 (2019).
22. K. Wang, L. Song, C. Wang, *et al.*, "On the use of deep learning for phase recovery," Light Sci. Appl. **13**, 4 (2024).
23. R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *International Conference on Learning Representations* (2019).
24. X. Tang, X. Zhao, J. Liu, *et al.*, "Uncertainty-aware unsupervised image deblurring with deep residual prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), pp. 9883–9892.
25. W. Li, K. Abrashitova, and L. V. Amitonova, "Super-resolution multimode fiber imaging with an untrained neural network," Opt. Lett. **48**, 3363–3366 (2023).
26. L.-H. Yeh, J. Dong, J. Zhong, *et al.*, "Experimental robustness of Fourier ptychography phase retrieval algorithms," Opt. Express **23**, 33214–33240 (2015).
27. G.-Z. Yang, B.-Z. Dong, B.-Y. Gu, *et al.*, "Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: a comparison," Appl. Opt. **33**, 209–218 (1994).
28. T. E. Gureyev, T. J. Davis, A. Pogany, *et al.*, "Optical phase retrieval by use of first Born- and Rytov-type approximations," Appl. Opt. **43**, 2418–2430 (2004).

29. B. Blaschke-Kaltenbacher and H. W. Engl, *Regularization Methods for Nonlinear Ill-Posed Problems with Applications to Phase Reconstruction* (Springer, 1997), pp. 17–35.

30. H. Egger and A. Leitão, "Nonlinear regularization methods for ill-posed problems with piecewise constant or strongly varying solutions," Inverse Probl. **25**, 115014 (2009).

31. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (Springer International Publishing, 2015), pp. 234–241.

32. C. Osorio Quero, D. Durini, J. Rangel-Magdaleno, *et al.*, "Single-pixel near-infrared 3D image reconstruction in outdoor conditions," Micromachines **13**, 795 (2022).

33. X. Qin, Z. Zhang, C. Huang, *et al.*, "U$^2$-Net: Going deeper with nested u-structure for salient object detection," Pattern Recognit. **106**, 107404 (2020).

34. D. Mery, V. Riffo, U. Zscherpel, *et al.*, "GDXray: The database of X-ray images for nondestructive testing," J. Nondestruct. Eval. **34**, 42 (2015).

35. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process. **21**, 4695–4708 (2012).

36. R. Fulton, S. Eberl, S. Meikle, *et al.*, "A practical 3D tomographic method for correcting patient head motion in clinical SPECT," IEEE Trans. Nucl. Sci. **46**, 667–672 (1999).

37. Z. Zhang, W. Sun, X. Min, *et al.*, "A no-reference visual quality metric for 3d color meshes," in *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (2021), pp. 1–6.

38. J. Dong, L. Valzania, A. Maillard, *et al.*, "Phase retrieval: from computational imaging to machine learning: a tutorial," IEEE Signal Process. Mag. **40**(1), 45–57 (2023).

39. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).

40. K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, "Phase retrieval from 1D Fourier measurements: convexity, uniqueness, and algorithms," IEEE Trans. Signal Process. **64**, 6105–6117 (2016).

41. W. Cong, N. Chen, and B. Gu, "A recursive method for phase retrieval in Fourier transform domain," Chin. Sci. Bull. **43**, 40–44 (1998).

42. G. Jagatap and C. Hegde, "Phase retrieval using untrained neural network priors," in *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks* (2019).

43. E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: theory and algorithms," IEEE Trans. Inf. Theory **61**, 1985–2007 (2015).

44. J. Ma, J. Xu, and A. Maleki, "Optimization-based AMP for phase retrieval: the impact of initialization and $\ell_2$ regularization," IEEE Trans. Inf. Theory **65**, 3600–3629 (2019).

45. M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **30**, 140–154 (1982).

46. D. Orsuti, C. Antonelli, A. Chiuso, *et al.*, "Deep learning-based phase retrieval scheme for minimum-phase signal recovery," J. Lightwave. Technol. **41**, 578–592 (2023).

47. E. Cha, C. Lee, M. Jang, *et al.*, "DeepPhaseCut: Deep relaxation in phase for unsupervised Fourier phase retrieval," IEEE Trans. Pattern Anal. Mach. Intell. **44**, 9931–9943 (2022).

48. L. Wu, S. Yoo, A. F. Suzana, *et al.*, "Three-dimensional coherent X-ray diffraction imaging via deep convolutional neural networks," npj Comput. Mater. **7**, 175 (2021).

49. Y. Zhang, M. A. Noack, P. Vagovic, *et al.*, "PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets," Opt. Express **29**, 19593–19604 (2021).

50. K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.

51. Y. Yang, L. Deng, P. Jiao, *et al.*, "Transfer learning in general lens-less imaging through scattering media," in *15th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (2020), pp. 1132–1141.

52. C.-J. Wang, C.-K. Wen, S.-H. Tsai, *et al.*, "Phase retrieval using expectation consistent signal recovery algorithm based on hypernetwork," IEEE Trans. Signal Process. **69**, 5770–5783 (2021).

53. D. Morales, A. Jerez, and H. Arguello, "Learning spectral initialization for phase retrieval via deep neural networks," Appl. Opt. **61**, F25–F33 (2022).

54. T. Uelwer, T. Hoffmann, and S. Harmeling, "Non-iterative phase retrieval with cascaded neural networks," in *Artificial Neural Networks and Machine Learning—ICANN 2021*, I. Farkaš, P. Masulli, S. Otte, and S. Wermter, eds. (Springer International Publishing, 2021), pp. 295–306.

55. Ç. Işil, F. S. Oktem, and A. Koç, "Deep iterative reconstruction for phase retrieval," Appl. Opt. **58**, 5422–5431 (2019).

56. W. Zuo, K. Zhang, and L. Zhang, *Convolutional Neural Networks for Image Denoising and Restoration* (Springer International Publishing, 2018), pp. 93–123.

57. F. Luo, J. Wang, J. Zeng, *et al.*, "Cascaded complex U-Net model to solve inverse scattering problems with phaseless-data in the complex domain," IEEE Trans. Antennas Propag. **70**, 6160–6170 (2022).

58. T. Zhang, "Phase retrieval by alternating minimization with random initialization," IEEE Trans. Inf. Theory **66**, 4563–4573 (2020).

59. F. Zhang, X. Liu, C. Guo, *et al.*, "Physics-based iterative projection complex neural network for phase retrieval in lensless microscopy imaging," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), pp. 10518–10526.

60. S. Pu, L. Li, Y. Xiang, *et al.*, "Phase retrieval based on enhanced generator conditional generative adversarial network," in *4th International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP)* (2022), pp. 825–829.

61. Y. Masuyama, K. Yatabe, Y. Koizumi, *et al.*, "Deep Griffin–Lim iteration: trainable iterative phase reconstruction using neural network," IEEE J. Sel. Top. Signal Process. **15**, 37–50 (2021).

62. P. Wen, M. Sun, and Y. Lei, "An improved U-Net method for sequence images segmentation," in *11th International Conference on Advanced Computational Intelligence (ICACI)* (2019), pp. 184–189.

63. S. Tozza and M. Falcone, "Analysis and approximation of some shape-from-shading models for non-Lambertian surfaces," J. Math. Imaging Vis. **55**, 153–178 (2016).

64. C. Y. Kao, S. Osher, and J. Qian, "Lax–Friedrichs sweeping scheme for static Hamilton–Jacobi equations," J. Comput. Phys. **196**, 367–391 (2004).

65. N. Siddique, S. Paheding, C. P. Elkin, *et al.*, "U-Net and its variants for medical image segmentation: a review of theory and applications," IEEE Access **9**, 82031–82057 (2021).

66. A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in *Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)* (2011), pp. 723–727.

67. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," IEEE Signal Process. Lett. **20**, 209–212 (2013).

68. C. J. Budd, A. T. McRae, and C. J. Cotter, "The scaling and skewness of optimally transported meshes on the sphere," J. Comput. Phys. **375**, 540–564 (2018).