



# PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets

YUHE ZHANG,<sup>1,\*</sup> MIKE ANDREAS NOACK,<sup>2</sup> PATRIK VAGOVIC,<sup>3</sup>   
KAMEL FEZZAA,<sup>4</sup> FRANCISCO GARCIA-MORENO,<sup>5</sup>   
RITSCHEL,<sup>6</sup> AND PABLO VILLANUEVA-PEREZ<sup>1</sup> 

<sup>1</sup>*Synchrotron Radiation Research and NanoLund, Lund University, Box 118, 221 00, Lund, Sweden*

<sup>2</sup>*Technische Universität Berlin, 10623 Berlin, Germany*

<sup>3</sup>*Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany*

<sup>4</sup>*X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA*

<sup>5</sup>*Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*

<sup>6</sup>*University College London, WC1E 6BT London, UK*

\*[yuhe.zhang@sljus.lu.se](mailto:yuhe.zhang@sljus.lu.se)

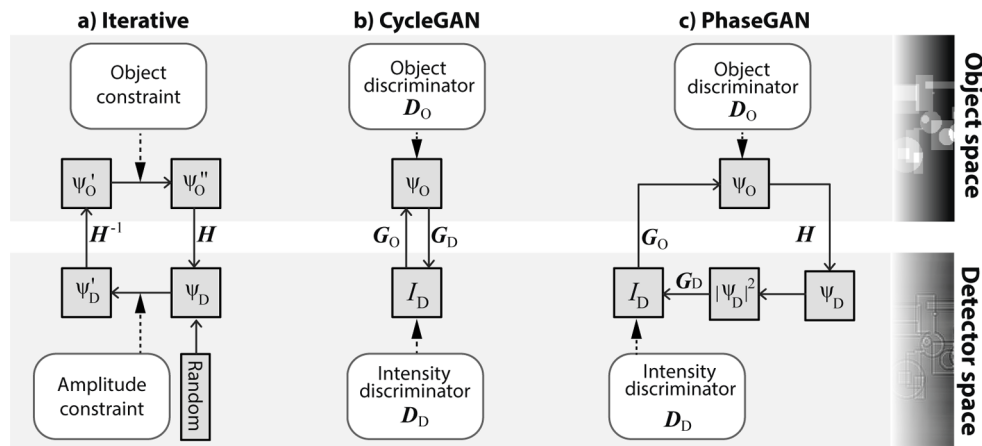
**Abstract:** Phase retrieval approaches based on deep learning (DL) provide a framework to obtain phase information from an intensity hologram or diffraction pattern in a robust manner and in real-time. However, current DL architectures applied to the phase problem rely on i) paired datasets, i. e., they are only applicable when a satisfactory solution of the phase problem has been found, and ii) the fact that most of them ignore the physics of the imaging process. Here, we present PhaseGAN, a new DL approach based on Generative Adversarial Networks, which allows the use of unpaired datasets and includes the physics of image formation. The performance of our approach is enhanced by including the image formation physics and a novel Fourier loss function, providing phase reconstructions when conventional phase retrieval algorithms fail, such as ultra-fast experiments. Thus, PhaseGAN offers the opportunity to address the phase problem in real-time when no phase reconstructions but good simulations or data from other experiments are available.

© 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Phase retrieval, i. e., reconstructing phase information from intensity measurements, is a common problem in coherent imaging techniques such as holography [1], coherent diffraction imaging [2], and ptychography [3,4]. As most detectors only record intensity information, the phase information is lost, making its reconstruction an ill-defined problem [5,6]. Most common quantitative solutions to the phase problem either rely on deterministic approaches or on an iterative solution [7]. Examples of deterministic solutions to holography are transport of intensity equations (TIE) [8] or based on Contrast Transfer Functions (CTFs) [9]. Such deterministic approaches can only be applied if certain constraints are met. For example, TIE is valid only in paraxial and short-propagation-distance conditions. Furthermore, complex objects can only be reconstructed with TIE when assuming a spatially homogeneous material [10]. Similarly, CTF only applies to weak scattering and weak absorption objects. Iterative approaches are not limited by these constraints [11,12] and can address not only holography but also coherent diffraction imaging and ptychography. These techniques retrieve the object by alternating between detector and object space and iteratively applying constraints on both, as depicted in Fig. 1(a). This process is computationally expensive, requiring several minutes to converge, precluding application to real-time analysis. Furthermore, the convergence of such approach is not guaranteed.

Recently, DL has demonstrated potential to solve ill-posed imaging problems, such as holography [13,14], magnetic resonance imaging [15], and phase retrieval [16–18]. DL offers an accurate solution to the phase problem, which is computationally fast compared to iterative



**Fig. 1.** Comparison between the schematic approaches of (a) conventional iterative phase-retrieval, (b) CycleGAN, and (c) PhaseGAN.

approaches [13,17], and independent of physical approximations. DL methods need to be *trained* before they are used. In *supervised* training, data is input to a differentiable method with adjustable parameters, e. g., a Neural Network (NN). The NN is then *tuned* to produce the desired output. In classic *paired* supervision applied to phase retrieval, for every input (intensity) such training needs to know the precise output (phase). Paired supervision has two difficulties:

First, such approaches require recording large datasets of phase and intensity of exactly the same sample. It is easy to think of conditions where this is not possible: i) Some instruments, like X-ray free-electron laser (XFEL) [19–22] have limited accessibility, making it difficult to acquire large paired datasets from such instruments. ii) Phase retrieval algorithms might not provide good reconstructions or are not even applicable. Examples of such scenarios are diffraction experiments where only simulations are available but not phase reconstructions [23,24] or Bragg Coherent Diffraction imaging [25] experiments where obtaining good phase reconstructions have proven a challenging task [26,27]. iii) Complementary imaging modalities, e. g., certain imaging experiments might provide low-noise and high-spatial-resolution phase reconstructions while another experiment provides high-noise detector images at a lower resolution of similar samples, but not of the same exact sample. This is of particular importance when imaging radio-sensitive samples with directly or indirectly-ionizing radiation, such as X-rays. This scenario requires minimizing the deposited dose, i. e., deposited energy per unit of mass. Alternatively, this is a typical problem when performing fast imaging experiments to track dynamics with a reduced number of photons per exposure. iv) Sensing might alter or even destroy the sample, e. g., in a diffraction-before-destruction imaging modality with high-intensity sources such as XFELs [28,29]. In this scenario, rendering paired sensing with a different modality is impossible. We argue how *unpaired* training, where all we need is random samples from the two different experimental setups, but not from the same object, will overcome all these four (i–iv) limitations.

Second, even if paired data was available, the results are often unsatisfying when attempting to solve an ill-posed problem, i. e., if one intensity reading does not map to one specific phase solution [30] but to a distribution of possible explanations. Classic paired training is known to *average*, i. e., spatially blur, all possible solutions if the solution is not unique [31]. Adversarial training [32] can overcome this problem by augmenting the training by a *discriminator*, i. e., another NN, with the purpose to correctly classify results from the training, as well as true samples of the data distribution, i. e., from-the-wild phase images, as either “real” or “fake”. The training uses the information of what was objectionable so that the discriminator could detect a

method's results as fake, to improve the method itself. It also uses the information from the true samples of the data distribution to become picky, i. e., good at saying what is “real” or “fake”. For ill-posed problems such as phase reconstruction, this will push the solution away from the average of all possible phase images to explain the input — which itself would not be a good phase image, as it is blurry — to a specific solution, which also explains the input, but is not blurry.

New DL adversarial schemes have shown the possibility of training on unpaired data sets; that is, a set of images captured from one modality and another set made using a different modality, but not necessarily of the same object. A state-of-the-art unpaired DL approach is CycleGAN [33]. CycleGAN learns a pair of *cycle consistent* functions, which map from one modality to the other such that their composition is the identity. This consistency constraint is analogous to the constraint applied in iterative phase reconstruction algorithms [5,11], where cyclic constraints are applied between the sample and detector space. Thus, approaches based on CycleGAN offer a framework for phase reconstruction, which mimics the structure of iterative approaches but without the limitation to paired datasets. For example, CycleGAN has been recently applied to holographic reconstructions [34] and medical imaging techniques including MRI [35], CT [36], and CT with super-resolution [37]. All these applications rely on the original CycleGAN framework, which does not include the physics of the image formation, and can lead to non-unique and non-robust solutions in the presence of noise [38].

Including physical knowledge into a neural network has been demonstrated to minimize error bounds [39]. This idea has been exploited together with deep-image prior [40] to address holographic reconstructions [41–43]. However, such approaches, similar to iterative approaches, require seconds to minutes to converge and provide phase reconstructions for each hologram. Thus, these approaches cannot be used to obtain real-time phase reconstructions [44] as required when performing ultra-fast or time-resolved experiments.

In this paper, we demonstrate a DL approach, christened PhaseGAN, based on CycleGAN. PhaseGAN includes the physics of the image formation as it cycles between the sample and the detector domains. By including the physics of the image formation and other learning constraints, PhaseGAN reconstructs the phase information at a level comparable to state-of-the-art paired approaches but obtaining better and more robust results than CycleGAN. Compared to current unsupervised DL approaches that include the physical model for phase reconstructions, PhaseGAN uses prior training. Thus, it can be used in real-time applications, such as time-resolved or ultra-fast imaging.

The remainder of this paper is structured as follows: First, we describe our approach's architecture and how the physics of the image formation is included. Second, we validate PhaseGAN with synthetic data for in-line holographic (near-field) experiments. In this validation step, we demonstrate the relevance of including the physical model by comparing the results with CycleGAN. Furthermore, we demonstrate that our unpaired approach performs at the level of state-of-the-art paired approaches. Third, we apply PhaseGAN to fast-imaging experimental data where noisy readings of a MHz camera are reconstructed using low-noise phase reconstructions recorded with a different setup and samples. Finally, we discuss the results and future applications of PhaseGAN to experiments where phase reconstructions are not possible today.

## 2. PhaseGAN approach

This section describes the architecture of PhaseGAN and how it uses physical knowledge to enhance the phase reconstructions. We then describe the training process and our loss function, which includes terms that avoid typical phase-reconstruction artifacts such as missing frequencies or the twin-imaging problem [1,45].

The architecture of PhaseGAN is based on CycleGAN [33]. CycleGAN uses two Generative Adversarial Network (GAN)s, which allow the translation of one image from a domain  $A$  to a

domain  $B$  and the inverse translation from  $B$  to  $A$ . Thus, the cycle consistency between two domains can be adapted to the object and detector domains, allowing CycleGAN to perform phase reconstructions by mimicking the structure of iterative phase-retrieval approaches, as shown in Fig. 1(b). The main difference between iterative phase-retrieval approaches and CycleGAN approaches is the inclusion of the propagator ( $H$ ), which includes the physics of the image formation between the object and the detector space. PhaseGAN combines both the iterative and the CycleGAN approach by including two GANs in a cyclic way together with the physics of the image formation via the propagator. The scheme of PhaseGAN is depicted in Fig. 1(c), where each of the GANs is decomposed in their generator ( $G$ ) and discriminator ( $D$ ). The generators used in PhaseGAN are U-Net [46]-like end-to-end fully convolutional neural networks. The discriminators are PatchGAN discriminators [31,33].  $G_O$  is the phase-reconstruction generator, which takes the measured intensities (one single channel input) and produces a two-channel complex output, where the two channels can be either the real and imaginary part or the phase and amplitude of the complex-object wave field ( $\psi_O$ ).  $D_O$  is the discriminator of the phase reconstruction. The object wavefield  $\psi_O$  is then propagated using the non-learnable but differentiable operator  $H$  to the detector plane ( $\psi_D = H\psi_O$ ), and the intensity in the detector plane is computed ( $|\psi_D|^2$ ). The propagator  $H$  is the near-field Fresnel propagator [47].  $G_D$  completes the cycle and works as an auxiliary generator, mapping the propagated intensity  $|\psi_D|^2$  to the measured detector intensity  $I_D = G_D|\psi_D|^2 = G_D|H\psi_O|^2$  using a single channel for the input and output. Due to the propagator  $H$ ,  $G_D$  does not need to learn the well-known physical process; thus it only learns the experimental effects of the intensity measurements, e. g., the point-spread function and flat-field artifacts. Although some of these effects can be described and included in the physical model such as the point-spread function, other effects cannot be included in a deterministic model, e. g., the shot noise due to the stochastic nature of the XFEL radiation [48]. Therefore,  $G_D$  learns all those experimental effects not included in the physical model. Finally, the intensity discriminator  $D_D$  is used to classify intensity measurements as “real” or “fake”. For more details about the PhaseGAN architecture, see the [Supplement 1](#).

Our goal is to learn two mappings simultaneously: i) detector images to complex object wavefield  $G_O : I_D \rightarrow \psi_O$ , and ii) propagated diffraction patterns to detector images  $G_D : |\psi_D|^2 \rightarrow I_D$ . This goal is achieved by optimizing

$$\arg \min_{G_O, G_D} \arg \max_{D_O, D_D} \mathcal{L}_{GAN}(G_O, G_D, D_O, D_D) + \alpha_{Cyc} \mathcal{L}_{Cyc}(G_O, G_D) + \alpha_{FRC} \mathcal{L}_{FRC}(G_O, G_D). \quad (1)$$

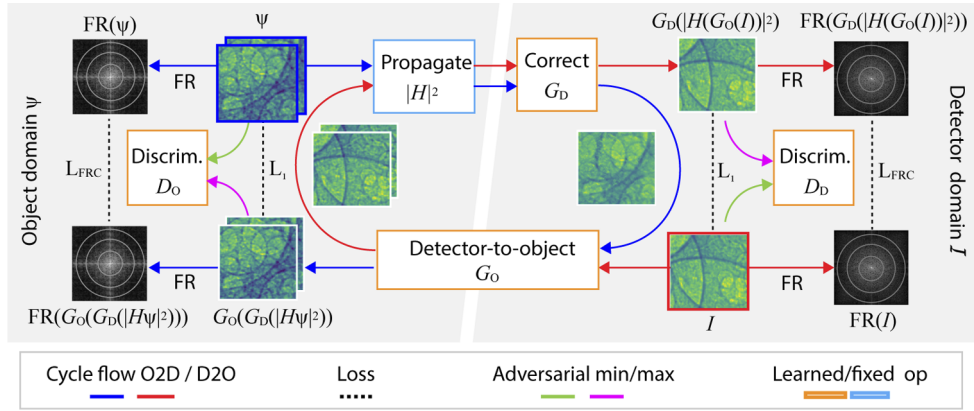
This objective is a combination of three terms: an adversarial term, a cycle consistency term, and a Fourier Ring Correlation (FRC) term. The relative weight of the cycle consistency and FRC losses with respect to the adversarial loss is parametrized by  $\alpha_{Cyc}$  and  $\alpha_{FRC}$ , respectively. The schematic of the learning process is depicted in Fig. 2.

The first term  $\mathcal{L}_{GAN}$  of Eq. (1) is the adversarial loss [32]

$$\begin{aligned} \mathcal{L}_{GAN}(G_O, G_D, D_O, D_D) = & \mathbb{E}_{\psi_O \sim \Psi} [\log(D_O(\psi_O))] + \\ & \mathbb{E}_{\psi_O \sim \Psi} [\log(1 - D_D(G_D|H\psi_O|^2))] + \\ & \mathbb{E}_{I_D \sim \mathcal{I}} [\log(D_D(I_D))] + \\ & \mathbb{E}_{I_D \sim \mathcal{I}} [\log(1 - D_O(G_O(I_D)))]. \end{aligned} \quad (2)$$

In Eq. (2),  $\mathbb{E}_{x \sim \mathcal{X}}$  denotes the expectation of the distribution  $\mathcal{X}$ , and  $\Psi$  and  $\mathcal{I}$  are the phase and intensity distributions, respectively.

The second term ( $\mathcal{L}_{Cyc}$ ) of Eq. (1) requires cycle consistency to confine generator outputs so that it is not just creating random permutation of images following the same data distribution



**Fig. 2.** Learning process diagram. Our aim is to learn a mapping  $G_O$  from the intensity sensing regime (right) to a phase modality (left). We require this mapping  $G_O$  to fulfill two cyclic constraints: First (blue), when its phase result is being mapped back to the intensity domain using a non-learned physical operator  $H$  and a learned correction operation  $G_D$ , the result should be similar (dotted line) to the intensity. Second (red), when the phase is mapped to intensity and back, it should remain the same. Further, we train two discriminators  $D_D$  and  $D_O$  to classify real and generated intensity and phase samples as real or fake (green). Finally, we ask the Fourier transform, another fixed but differentiable operator of both intensity and phase, to match the input after one cycle.

from the desired dataset. As shown in Fig. 2, regardless of where we start the loop we should end up at the starting point, i. e.,  $G_O(G_D|H\psi_O|^2) = \psi_O$  and  $G_D|H(G_O(I_D))|^2 = I_D$ . This cycle consistency loss can be expressed as:

$$\mathcal{L}_{\text{Cyc}}(G_O, G_D) = \mathbb{E}_{\psi_O \sim \Psi} [\|G_O(G_D|H\psi_O|^2) - \psi_O\|_1] + \mathbb{E}_{I_D \sim \mathcal{I}} [\|G_D|H(G_O(I_D))|^2 - I_D\|_1]. \quad (3)$$

The last term in Eq. (1),  $\mathcal{L}_{\text{FRC}}$ , calculates the FRC. FRC takes two images or complex waves and measures the normalised cross-correlation in Fourier space over rings [49,50]. Fourier ring correlation can help to avoid common frequency artifacts such as the twin-image problem [1,45] or missing frequencies due to the physical propagation. The  $\mathcal{L}_{\text{FRC}}$  is defined as follows:

$$\mathcal{L}_{\text{FRC}}(G_O, G_D) = \mathbb{E}_{\psi_O \sim \Psi} [\|1 - \text{FRC}(G_O(G_D|H\psi_O|^2), \psi_O)\|_2] + \mathbb{E}_{I_D \sim \mathcal{I}} [\|1 - \text{FRC}(G_D|H(G_O(I_D))|^2, I_D)\|_2], \quad (4)$$

where **FRC** is the Fourier ring correlation operator that calculates the FRC over all the Fourier space rings.

### 3. Validation results

In this section, we perform phase-retrieval experiments to validate PhaseGAN. Furthermore, we compare its performance to other state-of-the-art DL methods. This comparison is made with synthetic data in the near-field regime.

To validate PhaseGAN and compare its performance to other DL methods, we generated synthetic X-ray imaging experiments in the near-field regime using CelabA dataset [51]. The synthetic training dataset consists of 10,000 complex objects and 10,000 synthetic detector images. These sets are unpaired. However, paired solutions for the detector and object simulations are available for validation purposes and training state-of-the-art paired approaches. The wavelength used in these experiments is  $\lambda = 1 \text{ \AA}$ , and the pixel size in object space is constrained to  $1 \mu\text{m}$ .

The complex wavefront of the objects is given by their transmissivity. The transmissivity is emulated by multiplying the CelabA face images with a random thickness ( $t$ ), up to a maximum ( $t_{\max}$ ) of 1  $\mu\text{m}$ .

The complex wavefront after the object in the projection approximation is given by:

$$\psi_{\text{O}}(\mathbf{r}) = \psi_i \exp(jknt(\mathbf{r})), \quad (5)$$

where  $\psi_i$  is the illumination wavefront at the object plane,  $k = 2\pi/\lambda$  is the wavenumber,  $n$  is the complex refractive index,  $\mathbf{r}$  are the frame coordinates, and  $t(\mathbf{r})$  is the frame thickness map. The refractive index of gold at 12.4 keV is used for our simulation. Then, this wavefront is propagated to the detector ( $\mathbf{H}\psi_{\text{O}}$ ) using the near-field propagator. The near-field detector has an effective pixel-size of 1  $\mu\text{m}$  (equals to the sample-simulated pixel size) and is assumed to be 10 cm away from the sample. We also include flat-field noise, i. e., variable  $\psi_i$  for each frame. This flat-field noise is simulated with 15 elements of a basis extracted by Principal Component Analysis (PCA) from MHz-imaging data coming from the European XFEL [52]. Examples of the simulated holograms can be found in the [Supplement 1](#). We assume that the detector has photon counting capabilities; thus, the noise has Poissonian behaviour. The approximate number of photons simulated is  $6.6 \cdot 10^7$  per frame.

We compare the performance of PhaseGAN to four other DL methods. The first is a classic supervised learning approach using paired datasets and an  $L_2$  loss, as used by most current phase-retrieval approaches. The second uses the same architecture as before, but with additional adversarial terms as in `pix2pix` [53,54]. The global loss function in this `pix2pix` method is defined by:



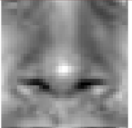
























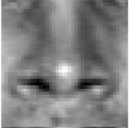











$$\begin{aligned} \mathcal{L}(\mathbf{G}_{\text{O}}, \mathbf{D}_{\text{O}}) = & \mathbb{E}_{\psi_{\text{O}} \sim \Psi} [\log(\mathbf{D}_{\text{O}}(\psi_{\text{O}}))] \\ & + \mathbb{E}_{\mathbf{I}_{\text{D}} \sim \mathcal{I}} [\log(1 - \mathbf{D}_{\text{O}}(\mathbf{G}_{\text{O}}(\mathbf{I}_{\text{D}})))] \\ & + \alpha_{\text{MSE}} \mathbb{E}_{(\psi_{\text{O}}, \mathbf{I}_{\text{D}}) \sim (\Psi, \mathcal{I})} \|\mathbf{G}_{\text{O}}(\mathbf{I}_{\text{D}}) - \psi_{\text{O}}\|_2. \end{aligned} \quad (6)$$

The first two terms of Eq. (6) calculate the adversarial loss in a similar way as we defined  $\mathcal{L}_{\text{GAN}}$  in Eq. (2). The weight of the  $L_2$  loss,  $\alpha_{\text{MSE}}$ , was set to 100. The third method is the standard `CycleGAN` approach presented in Fig. 1(b). We use the same global loss function as expressed in Eq. (1), but without including the physics of the image formation ( $\mathbf{H}$ ) as in Eqs. (2), (3), and (4). For the training of `CycleGAN`, we set  $\alpha_{\text{Cyc}} = 10$  and  $\alpha_{\text{FRC}} = 0$ . Last, we also report the results of PhaseGAN trained with  $\alpha_{\text{FRC}} = 0$ , which is denoted by PhaseGAN\*. Compared to PhaseGAN, PhaseGAN\* neglects the  $\mathcal{L}_{\text{FRC}}$  term in the optimization process. We set  $\alpha_{\text{Cyc}} = 10$  for the training of both PhaseGAN and PhaseGAN\*. For PhaseGAN training, we set  $\alpha_{\text{FRC}} = 5$ . For all experiments, we use the same phase-retrieval network  $\mathbf{G}_{\text{O}}$  and the same training dataset. The dataset was paired for the training of the first two methods but unpaired for the others. The ADAM optimizer [55] with a mini-batch size of 40 was used throughout the training. The generator learning rates were set to be 0.0002 for all five methods. For `pix2pix`, `CycleGAN`, PhaseGAN, and PhaseGAN\*, the discriminator learning rates were set to be 0.0001. We decayed all learning rates by 10 every 30 epochs and stopped training after 100 epochs.

The phase-retrieved results are quantified by using  $L_2$  norm, Dissimilarity Structure Similarity Index Metric (DSSIM) [56], and Fourier Ring Correlation Metric (FRCM). FRCM calculates the mean square of the difference between the Fourier ring correlation and unity over all spatial frequencies. Thus, smaller FRCM values imply a higher similarity between two images in the spectral or frequency domain. Please note that such metrics are only partially able to capture the ability of a GAN to produce data distribution samples [57]. It must also be considered that while these metrics assume the reference solution to be available, it is—for our method and `CycleGAN`—only used to compute the metric, never in training. For qualitative assessment, a reader is referred to Tbl. 1. Table 1 depicts the phase and amplitude of two zoom-in areas from the validation dataset and the corresponding reconstructions. Image patches shown in the same

column were plotted with the same color scale. In Tbl. 1, we also report, for each DL method, the logarithmic frequency distribution, the average value ( $\mu$ ), and the standard deviation ( $\sigma$ ) for the aforementioned validation metrics over 1000 validation images. Histograms in the same column were plotted using the same scale. More information about the statistical distribution of the metric values and line profiles through different validation images can be found in the [Supplement 1](#). Another validation with experimental data for the unpaired approaches with well-known objects is also presented in the [Supplement 1](#). The results there presented are consistent with the synthetic validation here presented. All the results presented here are discussed in section 5..

**Table 1. Comparison of different methods (rows) applied to the same input according to different metrics (columns).**

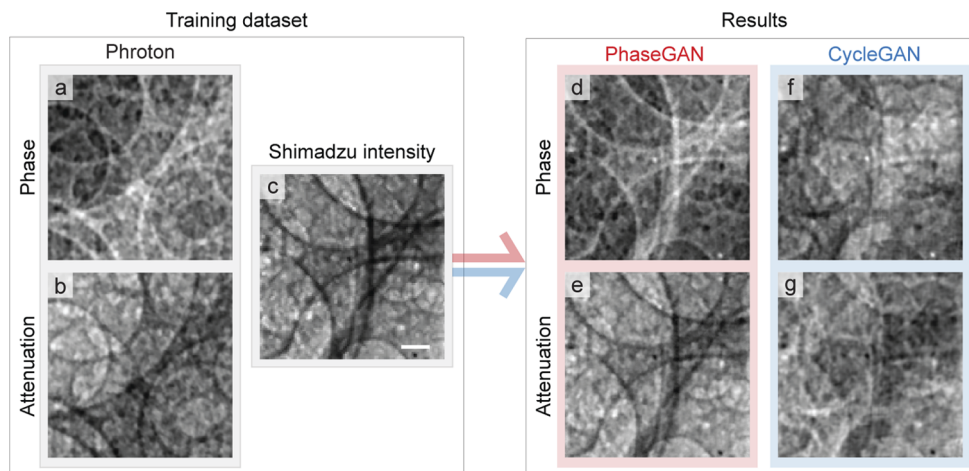
	Example patch 1		Example patch 2		$L_2$	DSSIM	FRCM
	Phase	Amplitude	Phase	Amplitude	$\times 10^{-4}$	$\times 10^{-3}$	$\times 10^{-2}$
oracle					—	—	—
paired					$\mu = 2.18$ $\sigma = 1.39$ 	$\mu = 6.57$ $\sigma = 2.91$ 	$\mu = 1.99$ $\sigma = 2.01$ 
pix2pix					$\mu = 1.89$ $\sigma = 1.15$ 	$\mu = 6.03$ $\sigma = 2.66$ 	$\mu = 1.84$ $\sigma = 1.79$ 
CycleGAN					$\mu = 43.3$ $\sigma = 22.5$ 	$\mu = 85.3$ $\sigma = 26.4$ 	$\mu = 31.7$ $\sigma = 4.1$ 
PhaseGAN*					$\mu = 7.74$ $\sigma = 4.47$ 	$\mu = 14.3$ $\sigma = 7.7$ 	$\mu = 2.86$ $\sigma = 2.91$ 
PhaseGAN					$\mu = 7.24$ $\sigma = 4.61$ 	$\mu = 12.4$ $\sigma = 7.3$ 	$\mu = 1.76$ $\sigma = 1.78$ 

#### 4. Experimental results

In this section, we applied PhaseGAN to experimental data recorded at the Advanced Photon Source (APS), where unpaired data of metallic foams was recorded with two different detectors at independent sensing experiments.

PhaseGAN offers the opportunity to obtain phase information when phase reconstructions are not possible. To demonstrate this, we performed time-resolved X-ray imaging experiments of the cell-wall rupture of metallic foams at the Advanced Photon Source (APS) [58]. The coalescence of two bubbles caused by the cell-wall rupture is a crucial process, which determines the final structure of a metallic foam [59]. This process can happen within microseconds; thus, MHz microscopic techniques are required to explore it. For this reason, we performed ultra-fast experiments with an X-ray imaging system based on a Photron FastcamSA-Z with 2  $\mu\text{m}$  effective

pixel size. The Photron system acquires the cell-wall rupture movies at a frame rate of 210 kHz, which integrated over 31 pulses of APS. Although the images acquired by the Photron camera used a few pulses, they had good contrast, which allows obtaining meaningful phase reconstructions. Images acquired by the Photron system were interpolated to an effective pixel size of  $1.6 \mu\text{m}$  and filtered using 100 iterations of a total-variation denoising algorithm [60] with denoising parameter  $\lambda = 1.5$ . Images obtained were phase-reconstructed using a TIE approach for single-phase materials [10] assuming X-ray photons of  $25.7 \text{ keV}$ ,  $\delta/\beta = 10^3$  and propagation distance  $z = 5 \text{ mm}$ . The reconstructed phase and attenuation for a frame of the Photron system are shown in Fig. 3(a) and (b), respectively. In order to increase the temporal resolution and to be able to use single pulses of APS, we used an X-ray MHz acquisition system based on a Shimadzu HPV-X2 camera with an effective pixel size of  $3.2 \mu\text{m}$ . This system was used to record movies of dynamic phenomena in liquid metallic foams using single pulses provided by APS with a repetition frequency of 6.5 MHz. An example of a frame recorded with this system is shown in Fig. 3(c). However, the contrast and signal-to-noise ratio are not sufficient to perform phase reconstructions with current approaches.



**Fig. 3.** Application to metallic foam. Phase (a) and attenuation (b) reconstructed using TIE of a frame acquired with the Photron-based system. (c) Intensity measured with the Shimadzu recording system using a single pulse. The length of the scale bar is  $50 \mu\text{m}$ . (d,e) Reconstructed phase and amplitude images of (c) obtained by PhaseGAN. (f,g) Corresponding phase and amplitude images of the same image retrieved by CycleGAN.

To overcome the impossibility of performing phase reconstructions using the frames recorded by the Shimadzu system, we used PhaseGAN. The dataset for PhaseGAN training consists of 10,000 Photron frames and 10,000 Shimadzu frames, with frame sizes of  $480 \times 200$  and  $128 \times 128$  pixels, respectively. Due to the different pixel sizes in the two imaging systems, the two sets of images were cropped to  $200 \times 200$  and  $100 \times 100$  before feeding them into the NN. This was done to match the field-of-view in the two different imaging domains. We performed data augmentation by applying random rotations and flips to the randomly cropped training images to take full advantage of PhaseGAN's capabilities. As is commonly used in supervised learning, data augmentation is also indispensable in unsupervised approaches for the neural network to learn the desired robustness properties [61], especially when only limited training examples are available. In our case, the holograms were captured by kHz to MHz camera systems, making detector frames very similar to each other. PhaseGAN reconstructions without data augmentation will not learn the desired mappings from one domain to the other but only remember the common features in each frame. The cropped Photron and Shimadzu



frames were subsequently padded during the training to  $256 \times 256$  and  $128 \times 128$ , respectively. We slightly modified the network architecture for the training of metallic foams, where an extra step of transposed convolution was added to the expanding step in  $G_O$  to double the size of the output images due to the half-pixel size of the Photron detector in respect of the Shimadzu one. Conversely, the last transposed convolutional layer of the  $G_D$  was replaced by a normal convolutional layer to accommodate the double-pixel size of the Shimadzu detector with respect to the Photron detector. We set  $\alpha_{Cyc} = 150$  and  $\alpha_{FRC} = 10$ . The ADAM optimizer with the same learning rates used for the synthetic data and a mini-batch size of 40 was adopted for the metallic foam training. The training was stopped after 100 epochs. The PhaseGAN phase and attenuation outputs for the Shimadzu frame depicted in Fig. 3(c) are shown in Fig. 3(d) and (e), respectively. The CycleGAN reconstructions of the same image using the same training parameters are shown in Fig. 3(f) and (g). A complete movie of the cell-wall rupture of a metallic foam (FORMGRIP alloy [62]) and its phase and attenuation reconstruction using PhaseGAN are provided in the supplemental [Visualization 1](#), [Visualization 2](#), and [Visualization 3](#). It is noticeable from the movie clip that the coalescence of the two bubbles was finished within  $10 \mu\text{s}$ . In total, 24.4 ms were consumed to reconstruct the 61 frames of the movie, i. e., PhaseGAN reconstructions took 0.4 ms per frame. Thus, PhaseGAN offers an opportunity for real-time analysis.

## 5. Discussion

We have presented PhaseGAN, a novel DL phase-retrieval approach. The cyclic structure of PhaseGAN allows to include the physics of image formation in the learning loop, which further enhances the capabilities of unpaired DL approaches, such as CycleGAN (see Tbl. 1). Although we did not include typical constraints used in iterative phase-retrieval approaches, such as support, histogram constraints, and sample symmetries, PhaseGAN performs at the level of state-of-the-art DL phase-reconstruction approaches. However, PhaseGAN's cyclic approach could be adapted to include such constraints to enhance its capabilities further. Another key ingredient of PhaseGAN is the inclusion of a FRC loss term, which penalizes common phase-reconstruction artifacts easy to filter in the Fourier domain, such as missing frequencies and the twin-imaging problem [1,45], and reconstructed better the details of the object. This enhancement can be seen in Tbl. 1 and [Supplement 1](#) when comparing the performance of PhaseGAN (with FRC loss term) and PhaseGAN\* (without FRC loss term).

We have demonstrated PhaseGAN's capabilities by performing near-field holographic experiments and compared the results to i) state-of-the-art paired approaches, ii) a GAN method following the pix2pix approach, and iii) CycleGAN. The results of the experiments, using the same training datasets, paired when needed, and phase-retrieval generator ( $G_O$ ), demonstrate the unique capabilities of PhaseGAN. These results are reported in Table 1. From this table, we can conclude that both paired approaches retrieve competitive phase reconstructions quantitatively and qualitatively. CycleGAN, due to the challenge of training on unpaired datasets, clearly performs worse than paired approaches. Without the inclusion of the physical model, CycleGAN can fail to distinguish the difference between phase and amplitude images and flips the order of the two channels, which increases the complexity of the learning process. PhaseGAN, although unpaired as well, retrieves results at the level of paired-training approaches and outperforms CycleGAN.

We have applied PhaseGAN to time-resolved X-ray imaging experiments using single pulses of a storage ring to study the cell-wall rupture of metallic foams. In this imaging modality, noisy images with low contrast and low resolution are recorded due to the limited number of photons per pulse. This acquisition scheme records images that cannot be phase-reconstructed. However, such an approach opens the possibility to record dynamics at MHz frame rates. In parallel, we acquired a less noisy and better-contrast dataset that allowed phase reconstructions. This dataset was obtained by integrating over 31 pulses and had about half of the pixel size of the

time-resolved dataset. By training using these two different sensing experiments on different realizations of metallic foam, we demonstrate the capability of PhaseGAN to produce phase reconstructions, which are not possible using paired approaches.

## 6. Conclusions

To conclude, we have presented a novel cyclic DL approach for phase reconstruction, called PhaseGAN. This approach includes the physics of image formation, a novel Fourier Ring Correlation loss function, and it can be used on unpaired training datasets to enhance the capabilities of current DL-based phase-retrieval approaches. We have demonstrated the unique capabilities of PhaseGAN to address the phase problem when no phase reconstructions are available, but good simulations of the object or data from other experiments are. This will enable phase reconstructions by correlating two independent experiments on similar samples. For example, it allows phase reconstructions and denoising with X-ray imaging from low-dose in-vivo measurements by correlating them with higher-dose and lower-noise measurements performed on ex-vivo samples of similar tissues and structures. As PhaseGAN uses prior training, it has the potential to denoise and reconstruct the phase of time-resolved experiments to track faster phenomena with a limited number of photons per frame. The capability of real-time phase-retrieval will be relevant for time-resolved experiments and ultra-fast imaging techniques.

The PhaseGAN code is available at [GitHub](#).

**Funding.** Bundesministerium für Bildung und Forschung (05K18KTA); Vetenskapsrådet (2017-06719).

**Acknowledgments.** We are grateful to Z. Matej for his support and access to the GPU-computing cluster at MAX IV. The presented research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of a Quadro P4000 GPU used for this research.

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

**Supplemental document.** See [Supplement 1](#) for supporting content.

## References

1. D. Gabor, "A new microscopic principle," *Nature* **161**(4098), 777–778 (1948).
2. J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," *Nature* **400**(6742), 342–344 (1999).
3. R. Hegerl and W. Hoppe, "Dynamische Theorie der Kristallstrukturanalyse durch Elektronenbeugung im inhomogenen Primärstrahlwellenfeld, Berichte der Bunsengesellschaft für physikalische Chemie (1970).," *Nature* **74**, 1148–1154 (1948).
4. H. M. Faulkner and J. M. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," *Phys. Rev. Lett.* **93**(2), 023903 (2004).
5. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**(15), 2758–2769 (1982).
6. M. R. Teague, "Irradiance moments: their propagation and use for unique retrieval of phase," *J. Opt. Soc. Am.* **72**(9), 1199–1209 (1982).
7. C. Zuo, J. Li, J. Sun, Y. Fan, J. Zhang, L. Lu, R. Zhang, B. Wang, L. Huang, and Q. Chen, "Transport of intensity equation: a tutorial," *Opt. Lasers Eng.* **135**, 106187 (2020).
8. M. Reed Teague, "Deterministic phase retrieval: a Green's function solution," *J. Opt. Soc. Am.* **73**(11), 1434 (1983).
9. J.-P. Guigay, "Fourier-transform analysis of Fresnel diffraction patterns and in-line holograms," *Optik* **49**, 121–125 (1977).
10. D. Paganin, S. C. Mayo, T. E. Gureyev, P. R. Miller, and S. W. Wilkins, "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object," *J. Microsc.* **206**(1), 33–40 (2002).
11. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**, 237–246 (1972).
12. J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.* **3**(1), 27 (1978).
13. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Sci. Appl.* **7**(2), 17141 (2018).
14. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Express* **26**(18), 22603 (2018).
15. T. Leiner, D. Rueckert, A. Suinesiaputra, B. Baeßler, R. Nezafat, I. Išgum, and A. A. Young, "Machine learning in cardiovascular magnetic resonance: Basic concepts and applications," *J. Cardiovasc Magn. Reson.* **21**(1), 61 (2019).

16. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**(9), 1117 (2017).
17. M. J. Cherukara, Y. S. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," *Sci. Rep.* **8**(1), 16520–16528 (2018).
18. Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica* **6**(5), 618–629 (2019).
19. J. M. Madey, "Stimulated Emission of Bremsstrahlung in a Periodic Magnetic Field," *J. Appl. Phys.* **42**(5), 1906–1913 (1971).
20. A. Kondratenko and E. Saldin, "Generating of coherent radiation by a relativistic electron beam in an undulator," *Part. Accel.* **10**, 207–216 (1980).
21. R. Bonifacio and F. Casagrande, "Classical and quantum treatment of amplifier and superradiant free-electron laser dynamics," *J. Opt. Soc. Am. B* **2**(1), 250–258 (1985).
22. R. Bonifacio, L. De Salvo, P. Pierini, N. Piovela, and C. Pellegrini, "Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise," *Phys. Rev. Lett.* **73**(1), 70–73 (1994).
23. A. Davtyan, S. Lehmann, D. Kriegner, R. R. Zamani, K. A. Dick, D. Bahrami, A. Al-Hassan, S. J. Leake, U. Pietsch, and V. Holý, "Characterization of individual stacking faults in a wurtzite GaAs nanowire by nanobeam X-ray diffraction," *J. Synchrotron Radiat.* **24**(5), 981–990 (2017).
24. A. Diaz, V. Chamard, C. Mocuta, R. Magalhães-Paniago, J. Stangl, D. Carbone, T. H. Metzger, and G. Bauer, "Imaging the displacement field within epitaxial nanostructures by coherent diffraction: a feasibility study," *New J. Phys.* **12**(3), 035006 (2010).
25. I. Robinson and R. Harder, "Coherent x-ray diffraction imaging of strain at the nanoscale," *Nat. Mater.* **8**(4), 291–298 (2009).
26. J. Carnis, L. Gao, S. Labat, Y. Y. Kim, J. Hofmann, S. Leake, T. Schüllli, E. Hensen, O. Thomas, and M.-I. Richard, "Towards a quantitative determination of strain in bragg coherent x-ray diffraction imaging: artefacts and sign convention in reconstructions," *Sci. Rep.* **9**(1), 17357 (2019).
27. Z. Wang, O. Gorobtsov, and A. Singer, "An algorithm for bragg coherent x-ray diffractive imaging of highly strained nanocrystals," *New J. Phys.* **22**(1), 013021 (2020).
28. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, "Potential for biomolecular imaging with femtosecond X-ray pulses," *Nature* **406**(6797), 752–757 (2000).
29. H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, R. A. London, E. Plönjes, M. Kuhlmann, R. Treusch, S. Düsterer, T. Tschentscher, J. R. Schneider, E. Spiller, T. Möller, C. Bostedt, M. Hoener, D. A. Shapiro, K. O. Hodgson, D. van der Spoel, F. Burmeister, M. Bergh, C. Caleman, G. Huldt, M. M. Seibert, F. R. N. C. Maia, R. W. Lee, A. Szöke, N. Timneanu, and J. Hajdu, "Femtosecond diffractive imaging with a soft-X-ray free-electron laser," *Nat. Phys.* **2**(12), 839–843 (2006).
30. Y. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," *Opt. Commun.* **30**(3), 304–308 (1979).
31. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. Computer Vision and Pattern Recognition*, (2017), pp. 4681–4690.
32. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Neural Information Processing Systems*, (2014), pp. 2672–2680.
33. J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), pp. 2242–2251.
34. D. Yin, Z. Gu, Y. Zhang, F. Gu, S. Nie, J. Ma, and C. Yuan, "Digital holographic reconstruction based on deep learning framework with unpaired data," *IEEE Photonics J.* **12**(2), 1–12 (2020).
35. T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss," *IEEE transactions on medical imaging* **37**(6), 1488–1497 (2018).
36. E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, "Cycle-consistent adversarial denoising network for multiphase coronary ct angiography," *Med. Phys.* **46**(2), 550–562 (2019).
37. C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, and G. Wang, "Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle)," *IEEE transactions on medical imaging* **39**(1), 188–203 (2020).
38. D. Bashkurova, B. Usman, and K. Saenko, "Adversarial self-defense for cycle-consistent gans," arXiv preprint arXiv:1908.01517 (2019).
39. A. Maier, F. Schebesch, C. Syben, T. Würfl, S. Steidl, J.-H. Choi, and R. Fahrig, "Precision learning: towards use of known operators in neural networks," in *2018 24th International Conference on Pattern Recognition (ICPR)* (IEEE, 2018), pp. 183–188..
40. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 9446–9454.
41. R. Hyder, V. Shah, C. Hegde, and M. S. Asif, "Alternating phase projected gradient descent with generative priors for solving compressive phase retrieval," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2019), pp. 7705–7709.

42. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light: Sci. Appl.* **9**(1), 77–87 (2020).
43. E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, "Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network," *Optica* **7**(6), 559–562 (2020).
44. R. Harder, "Deep neural networks in real-time coherent diffraction imaging," *IUCrJ* **8**(1), 1–3 (2021).
45. T. Latychevskaia and H.-W. Fink, "Solution to the twin image problem in holography," *Phys. Rev. Lett.* **98**(23), 233901 (2007).
46. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, (2015), pp. 234–241.
47. M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Cambridge University Press, 1999), 7th ed.
48. R. Bonifacio, L. De Salvo, P. Pierini, N. Piovella, and C. Pellegrini, "Spectrum, temporal structure, and fluctuations in a high-gain free-electron laser starting from noise," *Phys. Rev. Lett.* **73**(1), 70–73 (1994).
49. W. Saxton and W. Baumeister, "The correlation averaging of a regularly arranged bacterial cell envelope protein," *J. Microsc.* **127**(2), 127–138 (1982).
50. M. van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *J. Struct. Biol.* **151**(3), 250–262 (2005).
51. Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, (2015).
52. P. Vagovič, T. Sato, L. Mikeš, G. Mills, R. Graceffa, F. Mattsson, P. Villanueva-Perez, A. Ershov, T. Faragó, J. Uličný, H. Kirkwood, R. Letrun, R. Mokso, M.-C. Zdora, M. P. Olbinado, A. Rack, T. Baumbach, J. Schulz, A. Meents, H. N. Chapman, and A. P. Mancuso, "Megahertz X-ray microscopy at X-ray free-electron laser and synchrotron sources," *Optica* **6**(9), 1106–1109 (2019).
53. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. Computer Vision and Pattern Recognition*, (2017), pp. 1125–1134.
54. X. Li, H. Qi, S. Jiang, P. Song, G. Zheng, and Y. Zhang, "Quantitative phase imaging via a cgan network with dual intensity images captured under centrosymmetric illumination," *Opt. Lett.* **44**(11), 2879–2882 (2019).
55. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
56. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Process.* **13**(4), 600–612 (2004).
57. Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. Computer Vision and Pattern Recognition*, (2018), pp. 6228–6237.
58. K. Fezzaa and Y. Wang, "Ultrafast x-ray phase-contrast imaging of the initial coalescence phase of two water droplets," *Phys. Rev. Lett.* **100**(10), 104501 (2008).
59. F. García-Moreno, M. Mukherjee, C. Jiménez, A. Rack, and J. Banhart, "Metal foaming investigated by x-ray radiography," *Metals* **2**(1), 10–21 (2011).
60. M. Lourakis, "TV-L1 image denoising algorithm," <https://www.mathworks.com/matlabcentral/fileexchange/57604-tv-l1-image-denoising-algorithm> (2020).
61. A. Dosovitskiy, J. T. Springenberg, M. Riedmiller, and T. Brox, "Discriminative unsupervised feature learning with convolutional neural networks," in *Proc. Neural Information Processing Systems*, (2014), pp. 766–774.
62. V. Gergely and B. Clyne, "The formgrip process: Foaming of reinforced metals by gas release in precursors," *Adv. Eng. Mater.* **2**(4), 175–178 (2000).