Applications of Neural Networks for the Detectability Enhancement in X-Ray Phase-Contrast Mammography

ZI-PING ZHONG[1] AND CHENG-YING CHOU[1,*]

[1]*Department of Biomechatronics Engineering, National Taiwan University, 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan 106*

[*]*chengying@ntu.edu.tw*

Abstract: X-ray phase-contrast imaging methods can reflect both the absorption and phase properties of an object. One of the most applications of propagation-based X-ray phase contrast is mammography. Phase retrieval algorithms can reconstruct phase and absorption images to reduce noise levels and dramatically enhance imaging contrast. However, an observer should measure the task-based image quality to guide the objective optimization of the imaging system. In this paper, the Hotelling observer (HO) is employed for signal detection tasks with signal-known-exactly tasks but with varying locations combined with background-known-statistically tasks to simulate the mammography structure. Furthermore, for mammogram inspection, this structure would be applied in a neural network to detect the signal to help doctors cope with diseases such as breast cancer or lesions more explicitly. Inspection in search of a lesion is a tough assignment that radiologists must perform frequently for weak-absorbing tissues, for example breast masses. Therefore, these methods are necessary to detect and delineate them, portraying crucial morphological information that will support a reliable diagnosis and is always inconspicuous for human observers. We use a generative adversarial network called "Enhanced PhaseGAN" to overcome this difficulty in this work. Afterward, receiver operating characteristic curves are utilized to evaluate the tractability and classification ability of the neural network using HO. The results indicate a significant improvement in image quality and level of contrast using a mixed approach optimization technique compared to conventional imaging.

1. Introduction

X-ray phase-contrast imaging can exploit an object's complex refractive index characteristics and reveal both the attenuation and phase properties of the imaged object. Phase-sensitive experimental setups such as interferometry-based [1] or analyzed-based [2] systems require sophisticated hardware design to meet the spatial or temporal requirements for phase-contrast imaging [3]. Propagation-based x-ray phase-contrast imaging (PB-PCI), on the other hand, can be realized by simply placing detector planes away from the object without the need for additional optics [4–7]. The phase shift after transmitting through an object can be obtained using a phase retrieval method [8].

Phase contrast is sensitive to electron density variation and can reveal structural details for objects with similar compositions. It is a promising and versatile imaging technique for many applications such as mammography [9], lung imaging [10], bone imaging [11, 12], micro-CT [13], and industrial nondestructive testing [14]. Recently, several studies have shown the performance of propagation-based phase-contrast imaging in mammography, which holds great potential to yield high-quality images with improved signal-to-noise ratio at high frequencies for screening mammography [15, 16].

In mammography, breast compression is used to reduce the thickness, but it causes the superimposition of lesions like micro-calcification and malignant cells on normal breast tissues.

To synthesize artificial mammograms, the background structures of breast tissues are usually simulated using a lumpy background model, which superimposes randomly positioned lumps. As a result, suspicious regions on mammograms can be obscured, preventing early detection of breast cancer. This paper uses clustered lumpy backgrounds (CLBs) to describe the background model more realistically [17]. Phase retrieval is an algorithmic process to extract the phase and attenuation information from intensity measurements. The retrieved phase data can provide images with increased contrast and information complementary to the conventional absorption images. Typical phase-retrieval methods such as contrast transfer function (CTF) and transport of intensity equation (TIE) are only valid under certain conditions [18]. CTF is valid under the assumption of a weakly absorbing object with smooth phase variation. TIE neglects the higher-order terms in the Taylor expansion of the transmittance function under the paraxial approximation and is only valid at short distances. The retrieved images will deviate from the actual projections when these assumptions are violated. The mixed approach between CTF and TIE is not limited to short propagation distances and produces better performance for strongly absorbing objects [19]. Despite the improvement gained by the mixed approach, the low-frequency noise caused by the zero-frequency singularity in the Fourier domain still cannot be eliminated. Suppose that a reconstruction algorithm can take into account the statistical properties of the underlying background and signal images. In that case, the low-frequency noise contamination can be mitigated to yield reconstructed images with enhanced contrast.

Recently, deep learning has been employed in tandem with conventional reconstruction algorithms to improve the reconstructed images in medicine [20]. Generative adversarial network (GAN) like CycleGAN was shown to realize the mapping from one domain to another for unpaired datasets [21], and it was utilized to remove the coherent noise in optical microscopy [22]. However, the image content cannot be preserved well since the physics of image formation, and inversion processes were not considered. Subsequently, PhaseGAN, based on CycleGAN, took into account the physics of the Fresnel operation and, at the same time, reduced the reconstruction error in the two domains [23]. It assumed materials with $\delta \sim 10^{-3}$ and $\beta \sim 10^{-6}$, which are significantly larger than the tissues observed in mammograms. However, the singularity in the reconstruction algorithm, which is attributed to lumpy noise, remains unaddressed because no perceptual similarity was taken into account to help recover the image features when the distorted effect is present. Thus, this paper proposes a new phase retrieval scheme to enhance signal detectability. Specifically, the covariance network is added to reduce the noise correlation [24] and the perceptual loss is included in the generators to capture the signal characteristics [25]. Thus, the features can be learned despite the presence of distortion. The proposed method is termed "Enhanced PhaseGAN" as it provides a framework for improving the phase reconstruction process and results in images with improved contrast and signal detectability.

The remainder of the paper is organized as follows. In Sec. 2, the physics of PB-PCI is briefly reviewed. The phase retrieval formulas such as CTF, TIE, and the mixed approach are presented in Sec. 3. In Sec. 4, the framework of the neural network-based models and our proposed method are described in detail. Subsequently, the training and validation of our proposed neural network-based method and the results of different methods are quantitatively compared in Sec. 5. Finally, conclusions are drawn in Sec. 7.

2. Imaging physics of inline phase-contrast X-ray imaging

Consider the schematic of the phase-contrast X-ray imaging shown in Fig. 1. An object can be characterized by its complex refractive index $n(\vec{r})$:

$$n(\vec{r}) = 1 - \delta(\vec{r}) + j\beta(\vec{r}), \tag{1}$$

where $\vec{r} \equiv (x, y, z)$ and $\delta(\vec{r})$ and $\beta(\vec{r})$ denote the X-ray refractive and absorption properties of the object. As a monochromatic X-ray wavefield U_i with the wavelength of λ travels along

the positive z-axis irradiated on the object, the transmitted wavefield is related to the incident wavefield as

$$U_o(x, y) = T(x, y)U_i, \tag{2}$$

where $T(x, y) = \exp\left[-A(x, y) + j\phi(x, y)\right]$ is the transmittance function of the object, and the projected absorption and phase shift are the linear integral of the imaginary and real parts of the complex-valued refractive index as

$$A(x, y) = k \int \beta(\vec{r})\, dz \tag{3}$$

and

$$\phi(x, y) = -k \int \delta(\vec{r})\, dz, \tag{4}$$

with $k = \frac{2\pi}{\lambda}$ denoting the wavenumber of the X-ray. The wavefield intensity recorded on the detector plane z_m downstream of the object can be described as

$$I_m(x, y) = |U_o(x, y) *_2 G_m(x, y)|^2 = |U_m(x, y)|^2, \tag{5}$$

where '$*_2$' denotes the two-dimensional (2D) convolution operation and G_m is the corresponding Fresnel operator. The projected phase and absorption contrasts can be obtained by applying a phase retrieval formula to intensity measurements taken at two or more distinct distances.
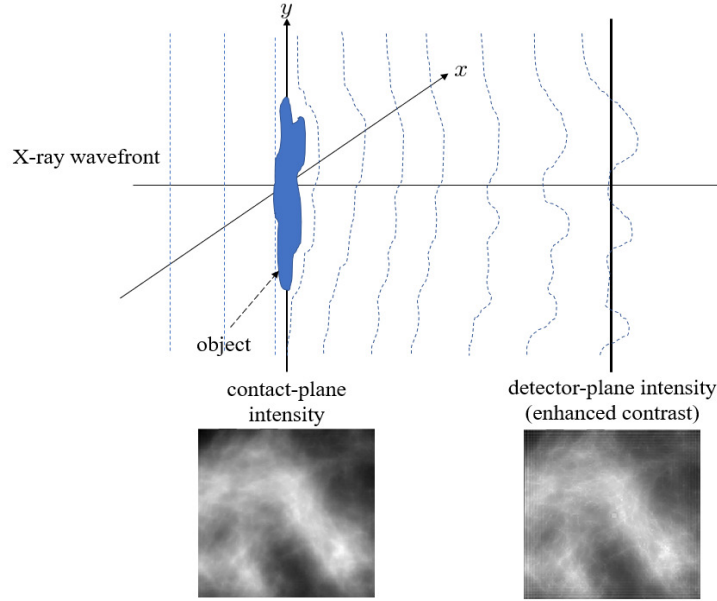

Fig. 1. The schematic diagram of propagation-based phase-contrast imaging.

3. Phase retrieval methods

Phase retrieval algorithms such as CTF and TIE are commonly used to quantitatively reconstruct phase and absorption images from the intensity measured at different detector planes. CTFs assume that the object is weakly absorbing with a smooth phase variation. TIE, on the other hand, neglects the higher-order terms in the Taylor expansion of the transmittance function under the paraxial approximation, which is only valid at short distances [18]. The mixed approach balanced

between the CTF and TIE can complement these two algorithms and breaks the underlying constraints. For the convenience of equation derivation, a modified data function is defined as

$$K_m(x, y) \equiv \frac{I_m(x, y)}{I_i} - 1, \tag{6}$$

where $I_i = |U_i|^2$ is the intensity of the incident X-ray beam. Here, the phase and absorption images derived from CTF [26] are

$$\widetilde{\phi}_{m,n}(u, v) = \frac{\cos\left(\pi\lambda z_n(u^2 + v^2)\right) \tilde{K}_m(u, v) - \cos\left(\pi\lambda z_m(u^2 + v^2)\right) \tilde{K}_n(u, v)}{D_{m,n}(u, v)} \tag{7}$$

and

$$\widetilde{A}_{m,n}(u, v) = \frac{\sin\left(\pi\lambda z_n(u^2 + v^2)\right) \tilde{K}_m(u, v) - \sin\left(\pi\lambda z_m(u^2 + v^2)\right) \tilde{K}_n(u, v)}{D_{m,n}(u, v)}, \tag{8}$$

where $\tilde{K}_m(u, v)$ is the Fourier transform of modified data function $K_m(x, y)$ as mentioned above. $D_{m,n}(u, v) \equiv 2\sin(\pi\lambda f^2(z_n - z_m))$, with subscripts m and n denoting the detector plane z_m and z_n. As the frequency f approaches zero, Eq. (7) and Eq. (8) become

$$\tilde{\phi}_{m,n}(u, v) = \frac{\tilde{K}_m(u, v) - \tilde{K}_n(u, v)}{2\pi\lambda f^2(z_n - z_m)} \tag{9}$$

and

$$\tilde{A}_{m,n}(u, v) = \frac{z_n\tilde{K}_m(u, v) - z_m\tilde{K}_n(u, v)}{2(z_n - z_m)}. \tag{10}$$

Note that there exists a singularity in Eq. (9) while none in Eq. (10). It can be seen that low-frequency noise is largely amplified in the phase image. On the other hand, its high-frequency components may also contain severe noise amplification if additional poles are presented. That is, when the term $D_{m,n}(u, v)$ in Eq. (7) satisfies $\lambda f^2(z_n - z_m) = \ell$ with ℓ equal to integers, these poles will emerge.

Thus, the major challenge is the ill-posed relationship between the intensities and the corresponding phase component due to these poles. It could be treated using the intensity spectrum under Wigner distribution [27], which is the linearized forward model. The phase shift derived based on this linearized forward model is as follows [18].

$$\tilde{\phi}(u, v) = \frac{1}{2B + \alpha} \times$$
$$\left[\sum_{i=1}^{N} \cos^2\left(\pi\lambda z_i(u^2 + v^2)\right) \sum_{i=1}^{N} \tilde{I}_i(u, v) \sin\left(\pi\lambda z_i(u^2 + v^2)\right) - \right. \tag{11}$$
$$\left. \sum_{i=1}^{N} \cos\left(\pi\lambda z_i(u^2 + v^2)\right) \sin\left(\pi\lambda z_i(u^2 + v^2)\right) \sum_{i=1}^{N} \tilde{I}_i(u, v) \cos\left(\pi\lambda z_i(u^2 + v^2)\right) \right],$$

where

$$B = \sum_{i=1}^{N} \cos^2\left(\pi\lambda z_i(u^2 + v^2)\right) \sum_{i=1}^{N} \sin^2\left(\pi\lambda z_i(u^2 + v^2)\right)$$
$$- \sum_{i=1}^{N} \cos^2\left(\pi\lambda z_i(u^2 + v^2)\right) \sin^2\left(\pi\lambda z_i(u^2 + v^2)\right), \tag{12}$$

and N is the number of distinct intensity measurements, which is 2 in this paper. α is applied as a regularization term introduced by the Tikhonov regularization to treat this low-frequency artifact by minimizing the variation in the background [18].

TIE describes the quantitative relationship between intensity and phase in wave propagation. It was derived under the paraxial and weak defocusing approximation, which requires intensity measurements to be taken at small distances. As mentioned in [18], the phase solution was derived as

$$\tilde{\phi}(u, v) = \mathscr{F}\left\{\frac{1}{2\pi\lambda z_m}\mathscr{F}^{-1}\left[\frac{1}{u^2 + v^2 + \alpha}\mathscr{F}\left(\frac{I_m(x, y)}{I_0(x, y)} - 1\right)\right]\right\}, \tag{13}$$

where $\mathscr{F}$ and $\mathscr{F}^{-1}$ denote Fourier and inverse Fourier transforms, respectively. A mixed approach was proposed by merging and extending the valid conditions of both methods [28]. Unlike the weak absorption assumption, the mixed transfer function model was derived under the slowly varying object approximation, which is more applicable as it allows for imaging strongly absorbing objects at a long propagation distance. To separate the phase information integrated in the intensity, a nonlinear iterative approach was utilized to obtain a more accurate solution of the phase. As shown in [27], the solution to the phase was estimated iteratively as

$$\tilde{\phi}^{(k)}(u, v) = \frac{\sum_{i=1}^{N} 2\sin\left(\pi\lambda z_i(u^2 + v^2)\right)\left(\tilde{I}_i^{(k)}(u, v) - \tilde{I}_i^{\phi=0}(u, v) - \Delta_i^{(k)}(u, v)\right)}{\sum_{i=1}^{N} 2\sin^2\left(\pi\lambda z_i(u^2 + v^2)\right)}, \tag{14}$$

where

$$\Delta_i^{(k)}(u, v) = \cos\left(\pi\lambda z_i f^2\right)\frac{\lambda z_i}{2\pi}\mathscr{F}\left\{\nabla \cdot [\phi^{(k-1)}(x, y)\nabla I_0(x, y)]\right\} \tag{15}$$

is a perturbation term, with the superscript (k) representing the k-th iteration. In our simulation trials, the phase estimate converges after 3 iterations. If a large propagation distance or high-resolution detector is employed, the denominator in Eq. (7) will lead to additional poles. Eq. (14), on the other hand, can eliminate the zero crossings caused by these additional poles when at least two distinct intensity measurements are employed [29]. However, even though intensity measurements at multiple distances are taken into account, the phase retrieval is still sensitive to noise at low spatial frequencies, which attribute to the lumpy appearance in phase images. Thus, α is applied to deal with it. According to [28], the initial guess of phase $\phi^{(0)}(x, y)$ is set to 0, hence $\Delta_i^{(1)}(u, v)$ becomes zero as well.

In the following sections, we use the reconstructed wavefield composed of phase and absorption images described in Eq. (8) and Eq. (14) as the input of the Enhanced PhaseGAN model.

4. Deep learning methods

Deep learning techniques can learn the underlying features and have the potential to complement conventional phase retrieval algorithms and generate phase images with improved image quality. CycleGAN combined adversarial losses and pixel-wise cycle consistency loss and could perform image-to-image translation between two domains for unpaired datasets [21]. To complete the phase reconstruction, PhaseGAN combined the physics of image formation described by the Fresnel propagator with CycleGAN [23]. However, simply including cycle consistency and adversarial losses fails to capture perceptual image structures in our experiment. To enhance the perceptual similarity and compensate for the lumpy appearance caused by the singularity in the phase retrieval algorithm, we propose a new deep learning-based phase retrieval framework termed Enhanced PhaseGAN, which is different from previous works in both network architecture and optimization design. In addition to two pairs of generators and discriminators in the object and detector domains, we introduce a covariance estimation network "CovNet", which can enhance image contrast and hence signal detectability. In the generators, Enhanced PhaseGAN combines adversarial loss and cycle consistency with perceptual loss based on a pretrained VGG model to enhance the quality of the output images.
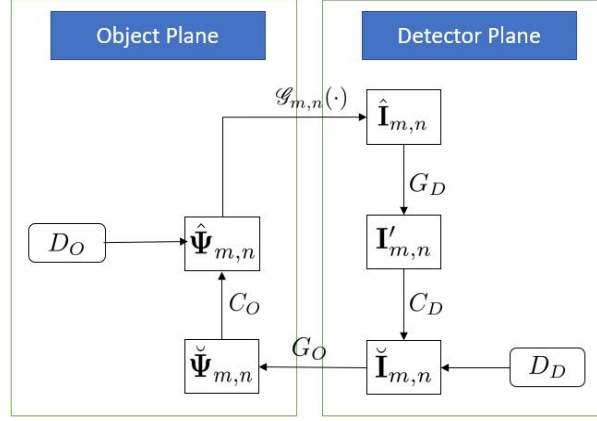
Fig. 2. The network design of Enhanced PhaseGAN.

4.1. The model architectures

The Enhanced PhaseGAN learns two mappings by translating images from one domain to the other, as shown in Fig. 2. It contains two domains: the object and detector planes, with each domain consisting of one pair of a generator and a discriminator. Here, the intensity vector $\mathbf{I}_{m,n}$ is taken as the input of the Enhanced PhaseGAN, denoting the intensity data on detector planes z_m and z_n:

$$\mathbf{I}_{m,n}^t = \left[I_m^t, \, I_n^t \right], \tag{16}$$

where the superscript t represents the transpose operation. On the other hand, the output is the reconstructed wavefield $\mathbf{\Psi}_{m,n}$, defined as

$$\mathbf{\Psi}_{m,n}^t = \left[\phi_{m,n}^t, \, A_{m,n}^t \right]. \tag{17}$$

Enhanced PhaseGAN aims to learn the relationship between two domains by using two pairs of generators and discriminators. The first generator is the detector-plane generator $G_D(\hat{\mathbf{I}}_{m,n}) = \mathbf{I}'_{m,n}$, which represents the image restoration process that reduces noise and recovers resolution loss for the intensity images on the detector-plane. The second is the object-plane generator $G_O(\breve{\mathbf{I}}_{m,n}) = \breve{\mathbf{\Psi}}_{m,n}$, which represents a phase retrieval process that maps the intensity distribution in the detector plane to phase estimates in the object plane. D_O and D_D represent the object-plane and detector-plane discriminators used to differentiate the generated images from the real ones.

As the off-diagonal elements of a covariance matrix represent the correlations among pixels, this model can help to reveal the pixel-wise correlations [30], which will further improve the predicted covariances for structure-preserving image denoising. In this paper, CovNet C_O and C_D in the object and detector planes are used to decrease the noise correlation to enhance the signal detectability of the reconstructed wavefields.

Phase and attenuation images are reconstructed by repeatedly feeding the wavefield and intensity into generators and covariance networks to learn and map the features and distributions between the two domains. To reduce loss and error in the training process, the discriminators measure the similarity between the ground truth and its reconstructed results.

The network architecture of the Enhanced PhaseGAN model is elaborated in the following subsections.

4.1.1. Discriminator

The discriminator is modified from the PatchGAN model, with the aim of restricting our attention to the image texture with considerable variation in local image patches and determining whether
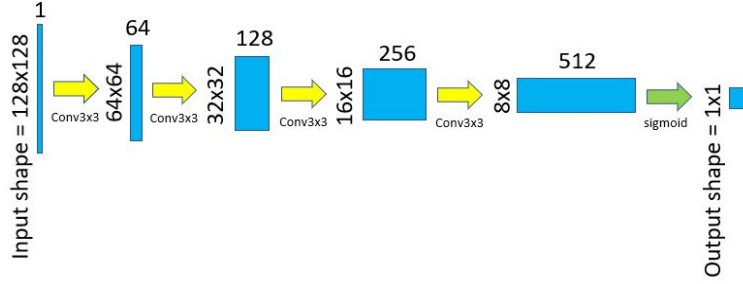
Fig. 3. Discriminator architecture.

each patch in an image is real or not. This model was proven to produce high-resolution images with improved details in the receptive field [31]. As shown in Fig. 3, its output is the average probability calculated using feature maps produced by the generators at different positions in the image. This probability can measure the fidelity of the image by scanning through different local patches within an image.

4.1.2. Generator

Three different kinds of network architectures, i.e., UNet, WNet, and SRResNet, are considered for the generators in the Enhanced PhaseGAN.

The UNet is designed for an image denoising task, which can treat the phase retrieval as a kind of inverse problem, and has been realized in GANs [32]. Its structure is shown in Fig. 4(a). For such tasks, the UNet here utilizes substitute layers in its extraction path, where the deconvolution layers are replaced with the upsampling operation. By doing so, the network can produce high-resolution images. Also, skip connection through copy and concatenation is applied among the layers in both the contraction and extraction paths to allow the subsequent layers to reuse the middle representations. After that, a "fully convolutional network" is applied to transform the image pixels into pixel categories for pixel-wise prediction, which is expected to produce salient signal features if present. Overall, the architecture consists of a contracting path to capture the context and a symmetric expanding path to reduce the reconstruction error, with precise signal localization if the signal is present.

The second kind of network we employed for the generators is the WNet, which is an autoencoder composed of two identical UNets [33], and its structure is shown in Fig. 4(b). The first UNet is an encoder for producing a pixel-wise prediction, while the second one is a decoder for reducing the reconstruction error. The encoder is a set of convolutional blocks and pooling operators that compress the input into compact features, while the decoder is a set of convolutional blocks and upsampling operators that are used to reconstruct the image through resizing the compact features generated from the encoder back to the original size of an image. Here, the objective of WNet is to check the correspondence of the images through an encoder that extracts critical features and information from the image. As for the decoder, it is to alleviate the sudden changes produced in the first part, and this arrangement enables further improved visualization with reduced artifacts in the image field.

The third one, SRResNet, is a resolution-based image model that aims to produce high-resolution images from those with low-resolution [34]. During the process, as shown in Fig. 4(c), skip-connection is used to connect the output feature maps of the last convolutional layer with the input feature maps of the residual blocks. Through that, it can prevent vanishing gradients during the training process.

4.1.3. Covariance estimation network: CovNet

We developed a covariance estimation network termed CovNet based on the structured uncertainty prediction network proposed by Dorta *et al.,* [24] to reduce noise correlation. This model can predict a full covariance matrix, which permits an efficient likelihood evaluation through capturing the pixel-wise correlations. Here, the CovNet is concatenated behind each generator in the Enhanced PhaseGAN to calculate the covariance matrix of the output of each generator.

The network structure shown in Fig. 5 consists of a series of residual blocks, which are composed of convolutional layers, batch normalization layers, and ReLu activation layers. These layers are connected to predict the covariance matrix that is realized through padding zeros in its upper triangular part and multiplying it with its transpose. The target of this network is to produce images with reduced correlation, which correspond to $\mathbf{I}_{m,n}$ and $\mathbf{\Psi}_{m,n}$ on detector and object planes, respectively. Thus, if the covariance can be well estimated, the network can take advantage of the predicted covariance to capture the high-frequency structured image features and then improve the reconstruction.

4.2. Loss functions

The total loss function contains three terms: adversarial loss L_{GAN} to match the distribution of generated images to the data distribution in the target domain, cycle consistency loss $L_{Cyc,VGG}$ to prevent the learned mappings from contradicting each other, and the covariance loss L_{Cov} to further enhance the signal detectability by minimizing the noise correlation. Accordingly, the total loss function is expressed as

$$L = L_{GAN} + \gamma_1 L_{Cyc,VGG} + \gamma_2 L_{Cov}. \tag{18}$$

Here, the regularization terms corresponding to the cycle consistency and covariance losses are parameterized by γ_1 and γ_2, respectively. The objective of Enhanced PhaseGAN is to learn the forward and inverse mappings simultaneously through optimizing the total loss function

$$\arg\min_{G_O,G_D} \arg\max_{D_O,D_D} L. \tag{19}$$

4.2.1. Adversarial loss

The adversarial loss is aimed at making the distributions in two domains indistinguishable. It is realized through training the discriminator to maximize the probability of correctly classifying a given patch as real or fake. On the other hand, it enables the generator network to remove artifacts through mapping. The loss function is expressed as

$$
\begin{aligned}
L_{GAN} =& \mathbb{E}_{\hat{\mathbf{\Psi}}_{m,n} \sim \mathbf{\Psi}}[\log(D_O(\hat{\mathbf{\Psi}}_{m,n}))] \\
&+ \mathbb{E}_{\hat{\mathbf{\Psi}}_{m,n} \sim \mathbf{\Psi}}[\log(1 - D_O(G_O(\tilde{\mathbf{I}}_{m,n})))] \\
&+ \mathbb{E}_{\hat{\mathbf{I}}_{m,n} \sim \mathbf{I}_{m,n}}[\log(D_D(\tilde{\mathbf{I}}_{m,n}))] \\
&+ \mathbb{E}_{\hat{\mathbf{I}}_{m,n} \sim \mathbf{I}_{m,n}}[\log(1 - D_D(G_D(\hat{\mathbf{I}}_{m,n})))],
\end{aligned}
\tag{20}
$$

where $\log(\cdot)$ denotes the logarithmic function with base 10. $\mathbb{E}_{x \sim P(X)}$ denotes the expectation operation of component X based on the distribution $P(X)$.

4.2.2. Cycle consistency loss and VGG feature loss

Theoretically, an adversarial loss can learn the mapping between the outputs obtained from two generators and makes their distributions more similar. However, even though the adversarial loss function can check whether the distributions produced from the two generators contradict each other or not, the image textures produced from the two generators may not match because

of the nonrestricted mapping space. That is to say, by merely including adversarial loss, an output may not correspond to only one possible input, and thus the model cannot translate the images bilaterally. Hence, the consistency loss function utilizes L1 norm loss to reduce the space of possible mapping processes by enforcing forward and backward consistency to ensure the relationship of image translation between two domains. In an attempt to permit perceptual similarity, VGG feature loss, which utilizes pixel-wise loss, is also included to minimize the reconstruction error without paired images. The loss function is expressed as

$$L_{Cyc,VGG} = L_{Cyc,G_D} + L_{Cyc,G_O} + L_{VGG,G_D} + L_{VGG,G_O}, \tag{21}$$

where

$$L_{Cyc,G_O} = \mathbb{E}_{\hat{\mathbf{\Psi}}_{m,n} \sim \mathbf{\Psi}}(||G_O(\tilde{\mathbf{I}}_{m,n}) - \mathbf{\Psi}||_1) \tag{22a}$$

$$L_{Cyc,G_D} = \mathbb{E}_{\hat{\mathbf{I}}_{m,n} \sim \mathbf{I}_{m,n}}(||G_D(\hat{\mathbf{I}}_{m,n}) - \mathbf{I}_{m,n}||_1), \tag{22b}$$

and

$$L_{VGG,G_O} = \mathbb{E}_{\hat{\mathbf{\Psi}}_{m,n} \sim \mathbf{\Psi}}\left(\frac{1}{W_{i,j}H_{i,j}} \sum_{a=1}^{W_{i,j}} \sum_{b=1}^{H_{i,j}} VGG_{i,j}(G_O(\tilde{\mathbf{I}}_{m,n}))_{a,b} - VGG_{i,j}(\mathbf{\Psi})_{a,b}\right) \tag{23a}$$

$$L_{VGG,G_D} = \mathbb{E}_{\hat{\mathbf{I}}_{m,n} \sim \mathbf{I}_{m,n}}\left(\frac{1}{W_{i,j}H_{i,j}} \sum_{a=1}^{H_{i,j}} \sum_{b=1}^{W_{i,j}} VGG_{i,j}(G_D(\hat{\mathbf{I}}_{m,n}))_{a,b} - VGG_{i,j}(\mathbf{I}_{m,n})_{a,b}\right). \tag{23b}$$

Here, $W_{i,j}$ and $H_{i,j}$ indicate the dimensions of the respective feature maps within the VGG network. $VGG_{i,j}$ denotes the feature map obtained by the i-th convolution (after activation) before the j-th maxpooling layer within the VGG16 network [35].

4.2.3. Covariance loss

The algorithm aims to minimize the loss function to diminish the noise correlations from the autocovariance map analysis [36]. This network utilizes a diagonal Gaussian maximum negative log-likelihood approach whose predicted output is the lower triangular matrix. It is derived from the pixel-pixel noise covariance matrix, i.e., $(\hat{\mathbf{\Psi}}_{m,n} - \mathbf{\Psi})(\hat{\mathbf{\Psi}}_{m,n} - \mathbf{\Psi})^T$, which gives the full statistical description of the mapping uncertainty between the ground truth and generated images according to [30]. Here, the loss function is defined as

$$L_{Cov} = L_{Cov,G_O} + L_{Cov,G_D}, \tag{24}$$

where

$$L_{Cov,G_O} = -\log\left(|\hat{\Sigma}(G_O(\tilde{\mathbf{I}}_{m,n}))|\right) + (G_O(\tilde{\mathbf{I}}_{m,n}) - \mathbf{\Psi})^T \hat{\Sigma}^{-1}(G_O(\tilde{\mathbf{I}}_{m,n}))(G_O(\tilde{\mathbf{I}}_{m,n}) - \mathbf{\Psi}) \tag{25}$$

and

$$L_{Cov,G_D} = -\log\left(|\hat{\Sigma}(G_D(\hat{\mathbf{\Psi}}_{m,n}))|\right) + (G_D(\hat{\mathbf{I}}_{m,n}) - \mathbf{I}_{m,n})^T \hat{\Sigma}^{-1}(G_D(\hat{\mathbf{I}}_{m,n}))(G_D(\hat{\mathbf{I}}_{m,n}) - \mathbf{I}_{m,n}), \tag{26}$$

and $\Sigma^{-1}(\cdot)$ is defined as the inverse covariance matrix of image. As seen in Eq. (25) and Eq. (26), the matrix $\hat{\Sigma}^{-1}(\cdot)$ in its second term is computed by inverting the multiplication of two lower triangular matrices outputted by the network, which is $\hat{\Sigma} = M_x M_x^T$, where M_x is an image vector, with x denoting the output image.

5. Computer simulation

5.1. Data preparation and numerical mammograms

The dataset consists of $5,000$ pairs of mammographic intensity measurements, with each pair containing a signal absent and a signal present case, and the size of each of them set to be 128×128. For the $2,500$ pairs of signal-present case, both signal-known-exactly/background-known-statistically (SKE/BKS) and signal-known-statistically/background-known-statistically (SKS/BKS) signal detection tasks are equally included. The dataset was split into three subsets: $4,000$ pairs for training, 500 pairs for validation, and the other 500 pairs for testing, respectively.

The mammograms are synthesized based on the CLB object model mentioned in Sec. 1, with its parameters displayed in Table 1. Afterward, the intensity measurements are obtained by making use of Beer-Lambert law related to the refractive indices of the breast tissues and Fresnel propagation described by Eqs. (3)-(5). From these measurements, by employing Eqs. (11), (13) and (14), different estimates of phase shift can be obtained. An example of images reconstructed using different algorithms and the groundtruth are displayed in Fig. 6, and the performance metrics concerning PSNR, SSIM, and RMSE are listed in Table 2. It shows that the reconstructed phase using the mixed approach possesses higher contrast than that using CTF and TIE in weak absorption conditions. In addition, as the absorption becomes stronger, it can be seen that the mixed approach also yields better perceptual quality (which consists of PSNR and RMSE) because the weak absorption requirement for CTF is not met, and the TIE is highly sensitive to noise.

Table 1. Parameters for the CLB object model [17]. Note that L_x and L_y are the half-axes of each lump. σ_x and σ_y denote the width of each lump. α and β are adjustable parameters that affect the histogram analysis and the slope of power spectrum of CLB.

Variable	L_x	L_y	σ_x	σ_y	α	β
Value	5	2	12	12	2.1	0.5

Table 2. Performance metrics related to an example of reconstructed images displayed in Fig. 6.

Detection task	Phase retrieval algorithm	PSNR	SSIM	RMSE
SKE/BKS	CTF	19.56	0.28	0.63
	TIE	23.33	0.28	0.53
	Mixed approach	23.45	0.28	0.53
SKS/BKS	CTF	17.95	0.28	0.66
	TIE	21.55	0.29	0.56
	Mixed approach	21.84	0.28	0.53

5.2. Hyper-parameter setting

The hyper-parameter setting related to the Enhanced PhaseGAN is shown in Table 3. Note that the weights in all layers are initialized from a normal distribution. In addition, the optimizer applied in UNet and WNet is RMSprop because several experiments have proven that by using

this kind of optimizer, the solution for nonconvex problems can be obtained [37]. On the other hand, the Adam optimizer is suitable for training deep neural networks because the learning rate can be adjusted, with its parameters β_1 and β_2 set to be 0.9 and 0.999, respectively according to our several trials of these parameters both in the range from 0 to 1. As the training results get closer to the global minimum, the learning rate gets smaller correspondingly until the validation loss converges. Prior to feeding the training data into the Enhanced PhaseGAN, the iterative approach here utilizes the regularization technique from the solution described in Eq. (14), with α set to 0.0002 determined by utilizing 5,000 pairs of mammographic intensity measurements that aim to mitigate low-frequency artifacts. During the training process, the initial learning rates of the generator and discriminator are set to 0.002 and 0.001, respectively. Specifically, these learning rates are fixed in RMSprop optimizer and are decayed by 0.0001 every 10 epochs in Adam optimizer, and then both training are stopped after 200 epochs. To find the optimal proportion of the terms in Eq. (18), the regularizing terms (γ_1, γ_2) are set to (0.2,0.5), which is the best combination by several trials in the range from (0,0) to (1, 1). The training process is implemented through computer simulations, with the Pytorch version v1.6.0 and CUDA version 11.1. The code ran on the NVIDIA GeForce RTX 3080 Ti using the Pytorch-GPU version to accelerate the training process.

Table 3. Hyper-parameter settings determined by validation procedure of neural networks.

Architecture	UNet	WNet	SRResNet
Activation function	tanh+ReLu	ReLu	ReLu
batch size	28	28	28
Number of steps	1000	1000	1000
Optimizer	RMSProp	RMSProp	Adam
Initial learning rate of generator	0.002	0.002	0.002
Initial learning rate of discriminator	0.001	0.001	0.001
Dropout	no	no	yes
Epochs	200	200	200
Elapsed time	14 days	15 days	15 days

6. Results and discussion

6.1. Testing results

To mitigate the noise amplification due to singularity in the conventional phase retrieval formula, Enhanced phaseGAN is utilized, which considers the physics of image formation and the iterative approach proposed in the mixed approach to enhance the capability of CycleGAN. The testing results are obtained based on hyper-parameters selected by the validation process. The training and validation loss curves using different network structures are plotted in Fig. 7. In this figure, the solid and dashed curves correspond to the training and validation losses, respectively. It can be observed that the curves oscillate strongly in the initial epochs due to the lumpy appearance, which deters the distribution matching between the two generators. The losses drop dramatically as epochs proceed. In order to monitor the training process, the performance metrics, which

are root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), are used to measure and compare the perceptual quality of the reconstructed phase images with those obtained using the mixed approach [38] and the optimization of the neural networks with respect to their corresponding groundtruths. The performance indicators and AUC are shown in Table 1. It can be seen that the AUC generated by the Enhanced PhaseGAN is comprehensively improved. In particular, the SRResNet outperforms other networks, reaching an AUC of 82.6%. The enhanced detectability represents improved diagnosis accuracy from a clinical standpoint. On the other hand, the reconstruction error using WNet is the lowest among the network architectures in the SKE/BKS detection task, while the reconstruction error obtained using SRResNet is the lowest in the SKS/BKS detection task. Furthermore, contrast level is another criterion for measuring signal detectability in the field of phase contrast imaging. The phase reconstructed using WNet has a more similar contrast to the ground truth and a higher PSNR for the SKE/BKS detection task in the network. Furthermore, an example of the reconstructed phase using different network structures is shown in Fig. 8. It can be seen that the detailed features in the reconstructed images are well recovered among different networks for both SKE/BKS and SKS/BKS detection tasks. Moreover, with the reduced covariance values produced by CovNet, an enhanced contrast in the reconstructed images is observed.

As for the different generator networks between PhaseGAN and Enhanced PhaseGAN, according to our trails shown in Table 4, the overall performances using Enhanced PhaseGAN outperform those using PhaseGAN. Specifically, from the aspect of AUC value, the signal detectability using SRResNet as the generator model performs better than other generator models.

Because of the fact that the depth of SRResNet is the deepest, some subtle features can be recovered with higher resolution. On the other hand, from the perspective of visualization quality, the PSNR values are higher using Enhanced PhaseGAN than using PhaseGAN. Besides, the covariance loss is also applied to capture the features of the generator output and at the same time maintain the high frequency elements of phase images. Among Enhanced PhaseGAN generator models, WNet is ranked first in SKE/BKS detection task, whereas SRResNet performs the best in SKS/BKS detection task. As shown in Fig. 8, the images reconstructed with our network were successfully retrieved, with its first row performing better than its second row.

6.2.   Observer studies

The task-based detection performance is measured by implementing on the ideal observer study to guide the optimization of imaging systems using a test statistic. The receiver operating characteristic (ROC) curve is generated by varying the decision threshold over the test statistic values and then computing the true positive and the false-positive fraction. It is utilized as the figure of merit to measure the performance of a binary classification.

The objective of this observer is to maximize AUC through the likelihood ratio test, whose value is calculated using the Markov chain Monte Carlo method [39, 40]. As mentioned in Sec 5.1, 500 pairs of intensity measurements are utilized to evaluate the detection performance of an ideal observer corresponding to different kinds of networks. Their AUCs in terms of the reconstructed wavefield and intensity measurements are shown in Table 4. From this table, it can be observed that the AUC values in terms of $\phi_{1,2}$ can catch up with those of I_1 and I_2 by implementing Enhanced PhaseGAN compared to those merely applying mixed approach. This suggests that some information loss caused by the singularities contained in $\phi_{1,2}$ when implementing phase retrieval algorithm as presented in Eq. (7) is well decreased. Likewise, the AUC value concerning the cross information of wavefield $\phi_{1,2} + A_{1,2}$ can almost reach that of intensities $I_1 + I_2$. Despite the fact that SKE/BKS detection task is feasible in industry, this task is not practical in medicine because there are no exact information about the signal, such as its shape, position, amplitude, and so on. Contrary to SKE/BKS detection task, SKS/BKS detection task is a more viable case that its signal is produced randomly. Thus, the metrics in SKE/BKS

detection task perform better than those in SKS/BKS detection task. After implementing neural networks, all components concerning different metrics comprehensively improved.

Table 4. The qualitative results were associated with a mixed approach and different neural networks under two detection tasks.

Detection task	Metrics		Mixed approach	PhaseGAN	Enhanced PhaseGAN		
				UNet	UNet	WNet	SRResNet
SKE/BKS	AUC	$\phi_{1,2}$	0.580	0.631	0.778	0.749	**0.826**
		$A_{1,2}$	0.791	0.824	0.827	0.828	**0.832**
		$\phi_{1,2} + A_{1,2}$	0.895	0.903	0.901	0.904	**0.905**
		I_1	0.821	0.880	0.864	0.882	**0.889**
		I_2	0.877	0.893	0.884	0.8961	**0.8963**
		$I_1 + I_2$	0.901	0.927	0.922	0.9301	**0.9302**
	PSNR$(\phi, \phi_{1,2})$		17.03	25.1	23.04	**26.05**	20.3
	SSIM$(\phi, \phi_{1,2})$		0.401	0.601	0.625	0.654	**0.798**
	RMSE$(\phi, \phi_{1,2})$		0.652	0.523	0.533	**0.457**	0.521
SKS/BKS	AUC	$\phi_{1,2}$	0.501	0.583	0.592	0.601	**0.605**
		$A_{1,2}$	0.642	0.705	0.705	0.709	**0.710**
		$\phi_{1,2} + A_{1,2}$	0.705	0.814	0.824	0.8261	**0.8264**
		I_1	0.773	0.783	0.787	0.802	**0.814**
		I_2	0.787	0.826	0.839	0.849	**0.856**
		$I_1 + I_2$	0.866	0.886	0.886	0.893	**0.895**
	PSNR$(\phi, \phi_{1,2})$		12.58	17.2	17.05	18.17	**18.30**
	SSIM$(\phi, \phi_{1,2})$		0.403	0.519	**0.597**	0.589	0.544
	RMSE$(\phi, \phi_{1,2})$		0.782	0.700	0.712	0.698	**0.697**

7. Conclusion

A kind of generative adversarial network, called Enhanced PhaseGAN, is developed in this paper to deal with an inverse problem, which refers to the phase retrieval algorithm in this paper. As the qualitative results are shown, the AUC values using model structures of PhaseGAN and Enhanced PhaseGAN outperform those only using the mixed approach, suggesting that the classification ability is improved. Developing a generally applicable loss function for various imaging tasks is an important subject for future research as the proper weight between loss functions is task dependent. Among the model structures, Enhanced PhaseGAN possesses a better classification ability than PhaseGAN, with SRResNet performing the best even though the absorption becomes stronger. It is also worth noting that as the Table 4 shown, the level of contrast and reconstruction error are improved as well. Besides, despite the fact that $I_1 + I_2$ contains all information, to

make the reconstructed wavefields more realistic so that the discriminator can be deceived, the details of the intensity subsequently obtained from these wavefields by applying Fresnel operator improves as well. This process reflects on the AUC values after applying both PhaseGAN and Enhanced PhaseGAN, which escalates by approximately 3 percent in both SKE/BKS and SKS/BKS detection tasks compared to those only using the mixed approach.

This paper developed the Enhanced PhaseGAN to mitigate low-frequency noise due to singularities in conventional phase retrieval algorithms. The proposed method yields larger AUCs than other phase retrieval methods, including CTF-based, TIE-based, mixed approach, and PhaseGAN. SRResNet performs the best even with stronger absorption. It is also worth noting that the contrast and reconstruction error levels have also been improved, as shown in Table 4. Furthermore, although $I_1 + I_2$ contains all the information, in order to make the reconstructed wavefields more realistic and thus to fool the discriminator, the intensity details obtained from these wavefields by applying the Fresnel operator subsequently are also obtained Improve. This process mirrors the area under the curve (AUC) values after applying PhaseGAN and Augmented PhaseGAN, which increases about 3% for the SKE/BKS and SKS/BKS detection tasks compared to the hybrid-method-only task.

It is irrefutable that deep learning models are more robust when more data are included. However, the complexity of the images also needs to be considered, which affects the adaptability of these models. In our case, as signals and backgrounds become more diverse and complex in clinical use, the relationship between the reconstructed image and its projected image varies greatly, so it may not be possible to confirm the accuracy of the deep learning model. The adaptability largely depends on the prior information of the projected image. According to the recent findings in this paper, the use of an ideal observer will definitely enhance the detectability of the signal, which provides a figure of merit for evaluating and optimizing the imaging system, helping to improve the detection performance of small lesions. In conclusion, with optimized network design, this deep learning framework can be further applied to early clinical diagnosis.

Irrefutably, the deep learning models would be more robust when more data are included. However, the complexity of the image is also considered that affects the adaptability of these models. In our case, as the signal and backgrounds become more diverse and as complex as seen in clinical use, the relationships between the reconstructed image and its projection images vary drastically so that the adaptability of deep learning models may not be confirmed, which highly depends on the prior information of projection images. According to the recent results shown in this paper, the signal detectability is definitely enhanced using an ideal observer, which provides a figure-of-merit for use in assessing and optimizing imaging systems and is helpful to improve the detection performance of small lesions by incorporating deep learning models. In conclusion, with the optimal network design, this deep learning framework may be further applied in clinical early diagnosis.

References

1. F. Pfeiffer, J. Herzen, M. Willner, M. Chabior, S. Auweter, M. Reiser, and F. Bamberg, "Grating-based X-ray phase contrast for biomedical imaging applications," Zeitschrift für medizinische Physik **23**, 176–185 (2013).
2. H. N. Chapman and K. A. Nugent, "Coherent lensless X-ray imaging," Nat. photonics **4**, 833–839 (2010).
3. A. Snigirev, I. Snigireva, V. Kohn, S. Kuznetsov, and I. Schelokov, "On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation," Rev. scientific instruments **66**, 5486–5492 (1995).
4. P. Baran, S. Pacile, Y. I. Nesterets, S. Mayo, C. Dullin, D. Dreossi, F. Arfelli, D. Thompson, D. Lockie, M. McCormack *et al.*, "Optimization of propagation-based X-ray phase-contrast tomography for breast cancer imaging," Phys. Medicine & Biol. **62**, 2315 (2017).
5. M. Carrel, M. A. Beltran, V. L. Morales, N. Derlon, E. Morgenroth, R. Kaufmann, and M. Holzner, "Biofilm imaging in porous media by laboratory X-ray tomography: Combining a non-destructive contrast agent with propagation-based phase-contrast imaging tools," PloS one **12**, e0180374 (2017).
6. M. Töpperwien, R. Gradl, D. Keppeler, M. Vassholz, A. Meyer, R. Hessler, K. Achterhold, B. Gleich, M. Dierolf, F. Pfeiffer *et al.*, "Propagation-based phase-contrast X-ray tomography of cochlea using a compact synchrotron source," Sci. reports **8**, 1–12 (2018).

7. M. Krenkel, M. Töpperwien, C. Dullin, F. Alves, and T. Salditt, "Propagation-based phase-contrast tomography for high-resolution lung imaging with laboratory sources," AIP Adv. **6**, 035007 (2016).

8. R. Lohr, "Phase retrieval methods for polychromatic propagation-based phase-contrast x-ray imaging," Master's thesis, University of Waterloo (2019).

9. T. Michel, J. Rieger, G. Anton, F. Bayer, M. W. Beckmann, J. Durst, P. A. Fasching, W. Haas, A. Hartmann, G. Pelzer *et al.*, "On a dark-field signal generated by micrometer-sized calcifications in phase-contrast mammography," Phys. Medicine & Biol. **58**, 2713 (2013).

10. S. Wilkins, T. E. Gureyev, D. Gao, A. Pogany, and A. Stevenson, "Phase-contrast imaging using polychromatic hard X-rays," Nature **384**, 335–338 (1996).

11. H. Wen, E. E. Bennett, M. M. Hegedus, and S. Rapacchi, "Fourier X-ray scattering radiography yields bone structural information," Radiology **251**, 910–918 (2009).

12. H. Rougé-Labriet, S. Berujon, H. Mathieu, S. Bohic, B. Fayard, J.-N. Ravey, Y. Robert, P. Gaudin, and E. Brun, "X-ray phase contrast osteo-articular imaging: a pilot study on cadaveric human hands," Sci. reports **10**, 1–8 (2020).

13. A. Horng, J. Stroebel, T. Geith, S. Milz, A. Pacureanu, Y. Yang, P. Cloetens, G. Lovric, A. Mittone, A. Bravin *et al.*, "Multiscale x-ray phase contrast imaging of human cartilage for investigating osteoarthritis formation," J. Biomed. Sci. **28**, 1–14 (2021).

14. K. Suzuki and I. Haig, "Investigation of the application of phase contrast imaging using a point x-ray source to industrial non-destructive testing," Philos. Transactions Royal Soc. A: Math. Phys. Eng. Sci. **372**, 20130036 (2014).

15. L. Brombal, F. Arfelli, P. Delogu, S. Donato, G. Mettivier, K. Michielsen, P. Oliva, A. Taibi, I. Sechopoulos, R. Longo *et al.*, "Image quality comparison between a phase-contrast synchrotron radiation breast ct and a clinical breast ct: a phantom based study," Sci. reports **9**, 1–12 (2019).

16. Y. Ruiz-Gonzalez, M. Perez-Diaz, D. Martínez-Aguila, M. Diaz-Barreto, I. Fleitas, R. Mora-Machado, L. Rigon, G. Tromba, and P. Bregant, "Objective measurements of image quality in synchrotron radiation phase-contrast imaging versus digital mammography," Int. journal computer assisted radiology surgery **11**, 181–188 (2016).

17. F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," Opt. express **4**, 33–43 (1999).

18. M. Langer, P. Cloetens, J.-P. Guigay, and F. Peyrin, "Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography," Med. physics **35**, 4556–4566 (2008).

19. L. De Caro, F. Scattarella, C. Giannini, S. Tangaro, L. Rigon, R. Longo, and R. Bellotti, "Combined mixed approach algorithm for in-line phase-contrast X-ray imaging," Med. physics **37**, 3817–3827 (2010).

20. M. L. Terpstra, M. Maspero, F. d'Agata, B. Stemkens, M. P. Intven, J. J. Lagendijk, C. A. van den Berg, and R. H. Tijssen, "Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time mri-guided radiotherapy," Phys. Medicine & Biol. **65**, 155015 (2020).

21. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision,* (2017), pp. 2223–2232.

22. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," Light. Sci. & Appl. **8**, 1–8 (2019).

23. Y. Zhang, M. A. Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, and P. Villanueva-Perez, "Phasegan: A deep-learning phase-retrieval approach for unpaired datasets," arXiv preprint arXiv:2011.08660 (2020).

24. G. Dorta, S. Vicente, L. Agapito, N. D. Campbell, and I. Simpson, "Structured uncertainty prediction networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* (2018), pp. 5477–5485.

25. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (2017), pp. 4681–4690.

26. C.-Y. Chou, Y. Huang, D. Shi, and M. A. Anastasio, "Image reconstruction in quantitative x-ray phase-contrast imaging employing multiple measurements," Opt. express **15**, 10002–10025 (2007).

27. 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).

28. M. Langer, P. Cloetens, and F. Peyrin, "Regularization of phase retrieval with phase-attenuation duality prior for 3-d holotomography," IEEE Transactions on Image Process. **19**, 2428–2436 (2010).

29. S. Zabler, P. Cloetens, J.-P. Guigay, J. Baruchel, and M. Schlenker, "Optimization of phase contrast imaging using hard x-rays," Rev. Sci. Instruments **76**, 073705 (2005).

30. R. Keskitalo, M. Ashdown, P. Cabella, T. Kisner, T. Poutanen, R. Stompor, J. Bartlett, J. Borrill, C. Cantalupo, G. De Gasperis *et al.*, "Residual noise covariance for Planck low-resolution data analysis," Astron. & Astrophys. **522**, A94 (2010).

31. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (2017), pp. 1125–1134.

32. J. Chi, C. Wu, X. Yu, P. Ji, and H. Chu, "Single low-dose ct image denoising using a generative adversarial network with modified u-net generator and multi-level discriminator," IEEE Access **8**, 133470–133487 (2020).

33. X. Xia and B. Kulis, "W-net: A deep model for fully unsupervised image segmentation," arXiv preprint arXiv:1711.08506 (2017).

34. B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops,* (2017), pp. 136–144.

35. B. Kim, M. Han, H. Shim, and J. Baek, "A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose ct images," Med. physics **46**, 3906–3923 (2019).

36. C.-Y. Chou and M. A. Anastasio, "Influence of imaging geometry on noise texture in quantitative in-line x-ray phase-contrast imaging," Opt. express **17**, 14466–14480 (2009).

37. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

38. U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study," J. Comput. Commun. **7**, 8–18 (2019).

39. H. Zhang, *Signal detection in medical imaging* (The University of Arizona, 2001).

40. S. Park, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Ideal-observer performance under signal and background uncertainty," in *Biennial International Conference on Information Processing in Medical Imaging,* (Springer, 2003), pp. 342–353.
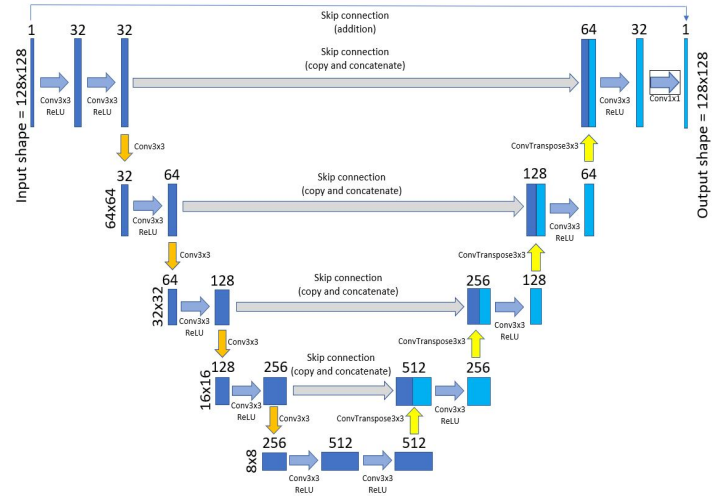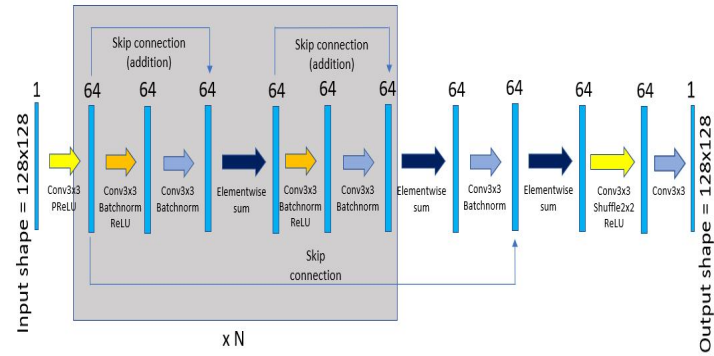
(a) UNet
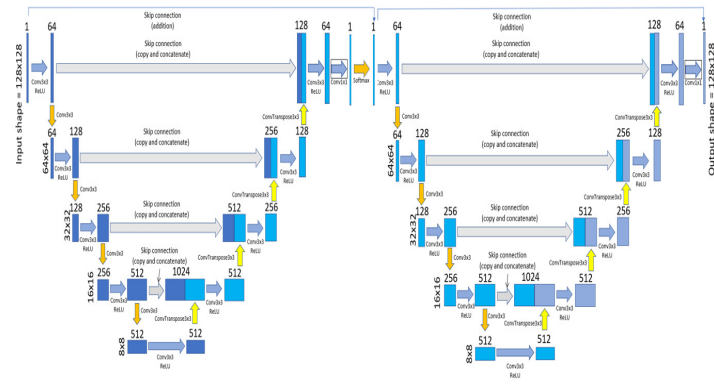
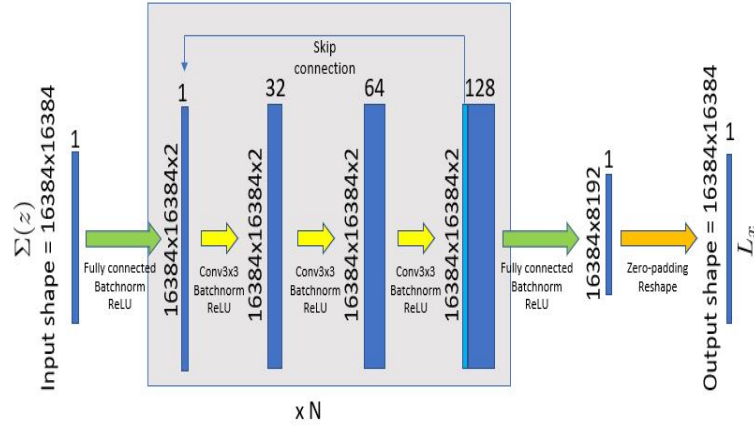
(b) WNet

(c) SRResNet

Fig. 4. Generator architectures.

Fig. 5. Covariance network: CovNet.

(a) $\phi_{\text{SKE/BKS}}$

(b) $\phi_{\text{SKS/BKS}}$

(c) $\phi_{1,2,\text{SKE/BKS}}^{\text{CTF}}$

(d) $\phi_{1,2,\text{SKE/BKS}}^{\text{TIE}}$

(e) $\phi_{1,2,\text{SKE/BKS}}^{\text{Mixed}}$

(f) $\phi_{1,2,\text{SKS/BKS}}^{\text{CTF}}$

(g) $\phi_{1,2,\text{SKS/BKS}}^{\text{TIE}}$
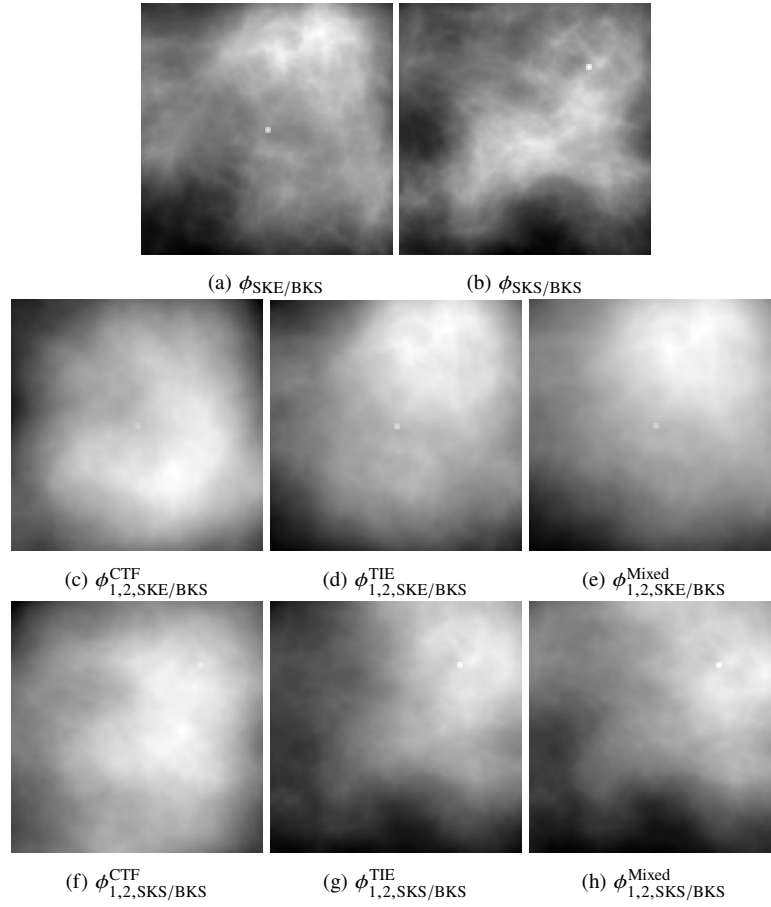
(h) $\phi_{1,2,\text{SKS/BKS}}^{\text{Mixed}}$

Fig. 6. The estimates of the projected phase and reconstructed phase in SKE/BKS (second row) and SKS/BKS (third row) detection tasks.
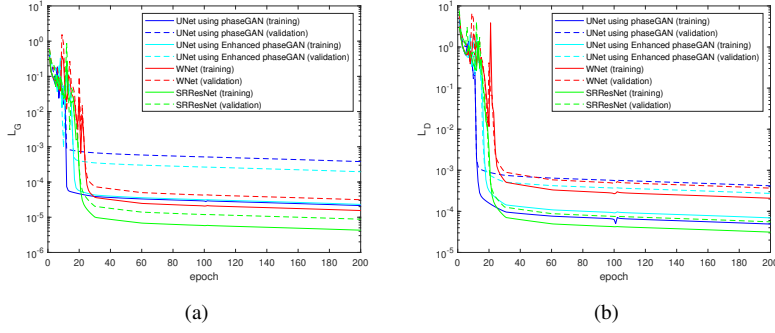
Fig. 7. The curves of learning process using loss functions in terms of (a) generator and (b) discriminator. Note that the curves displayed in solid line represent the training phase, while those displayed in dashed line represent the validation phase.
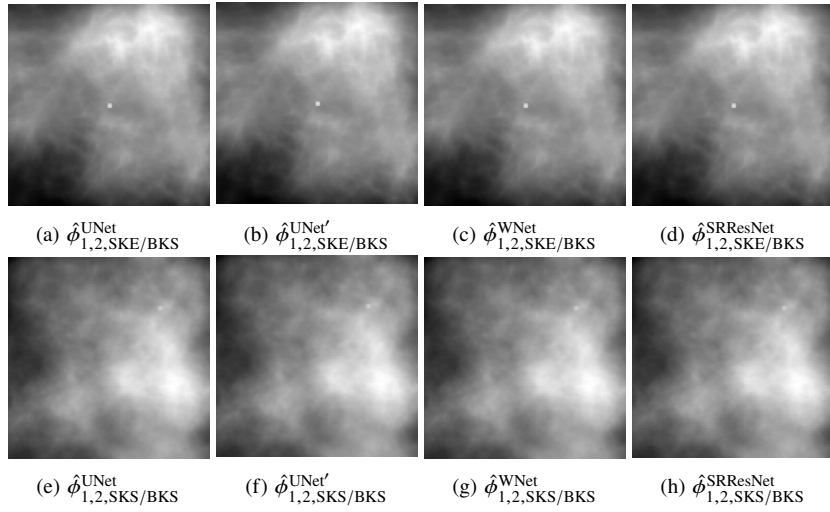

(a) $\hat{\phi}_{1,2,\text{SKE/BKS}}^{\text{UNet}}$ (b) $\hat{\phi}_{1,2,\text{SKE/BKS}}^{\text{UNet}'}$ (c) $\hat{\phi}_{1,2,\text{SKE/BKS}}^{\text{WNet}}$ (d) $\hat{\phi}_{1,2,\text{SKE/BKS}}^{\text{SRResNet}}$

(e) $\hat{\phi}_{1,2,\text{SKS/BKS}}^{\text{UNet}}$ (f) $\hat{\phi}_{1,2,\text{SKS/BKS}}^{\text{UNet}'}$ (g) $\hat{\phi}_{1,2,\text{SKS/BKS}}^{\text{WNet}}$ (h) $\hat{\phi}_{1,2,\text{SKS/BKS}}^{\text{SRResNet}}$

Fig. 8. The enhanced phase images with respect to UNet using phaseGAN, UNet using Enhanced phaseGAN (denoted as UNet′), WNet and SRResNet network structures from reconstructed phase images. Note that the images in the first row are reconstructed in SKE/BKS detection task using Fig. 6(e), whereas those displayed in the second row are in SKS/BKS detection task using Fig. 6(h).