Supplemental Document

Optics EXPRESS

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

YUHE ZHANG,[1,*] MIKE ANDREAS NOACK,[2] PATRIK VAGOVIC,[3] KAMEL FEZZAA,[4] FRANCISCO GARCIA-MORENO,[5] TOBIAS RITSCHEL,[6] AND PABLO VILLANUEVA-PEREZ[1]

[1] *Synchrotron Radiation Research and NanoLund, Lund University, Box 118, 221 00, Lund, Sweden*
[2] *Technische Universität Berlin, 10623 Berlin, Germany*
[3] *Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany*
[4] *X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA*
[5] *Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*
[6] *University College London, WC1E 6BT London, UK*
[*] *yuhe.zhang@sljus.lu.se*

Supplement DOI: https://doi.org/10.6084/m9.figshare.14686857

Parent Article DOI: https://doi.org/10.1364/OE.423222

PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets: Supplemental document

YUHE ZHANG,[1,*] MIKE ANDREAS NOACK,[2] PATRIK VAGOVIC,[3] KAMEL FEZZAA,[4] FRANCISCO GARCIA-MORENO,[5] TOBIAS RITSCHEL,[6] AND PABLO VILLANUEVA-PEREZ[1]

[1] *Synchrotron Radiation Research and NanoLund, Lund University, Box 118, 221 00, Lund, Sweden*
[2] *Technische Universität Berlin, 10623 Berlin, Germany*
[3] *Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany*
[4] *X-ray Science Division,Advanced Photon Source,Argonne National Laboratory, Lemont, IL 60439,US*
[5] *Helmholtz-Zentrum Berlin für Materialien und Energie, 14109 Berlin, Germany*
[6] *University College London, WC1E 6BT London, UK*
[*] *yuhe.zhang@sljus.lu.se*

Abstract: This document provides supplementary information to "PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets." In this material, we elaborate on the architecture of PhaseGAN. We also report and depict the results obtained by PhaseGAN when applied to the validation (synthetic) and experimental dataset. Finally, we present an experimental validation using well-known samples.

1. PhaseGAN Architecture

This section describes the architecture used for PhaseGAN.

The generators used in PhaseGAN are U-Net [1] type end-to-end fully convolutional neural networks. As shown in Fig. S1, the generator architecture consists of a contracting and expansive path. In the contracting path, the spatial resolution is reduced, and the feature information is increased. The contracting path in our model contains multiple convolutional layers with kernel size 3×3, each followed by a ReLU activation function. Max pooling operations with kernel size 2×2 are applied to 5 of the convolutional layers. After each max pooling, the image size is reduced by 2, decreasing from 256×256 to 8×8 pixels in the lowest resolution. The number of feature layers is doubled after each pooling operation. The extracted feature information is relocalized in the expansive path by combining upsampled feature mapping with the skip-connected high-resolution components from the contracting path. In the expansive path, the resolution of the images is recovered by repeated application of transposed convolutions. The transposed convolution outputs are then concatenated with the associated feature map from the contracting path and then send into corresponding convolutional layers. The generator weights are initialized by a pre-trained VGG11 encoder to improve model performance and accelerate the training process [2].

The discriminators used in this work are PatchGAN discriminators similar to the ones used in [3, 4]. They contain four convolutional layers with 4×4 convolutional filters, gradually increasing the number of filters by a factor of 2 from 64 to 512. Each convolution layer is followed by a batch normalization layer and a leaky RELU activation function with a slope of 0.2. The discriminators are trained to distinguish real images from the ones faked by the generator. For an image of size 256×256, the discriminator output a 30×30 matrix, where each matrix element corresponds to a 70×70 image area, examining if this part is from the training dataset or not.

The PhaseGAN architecture was trained using the MAX IV computing cluster. We used

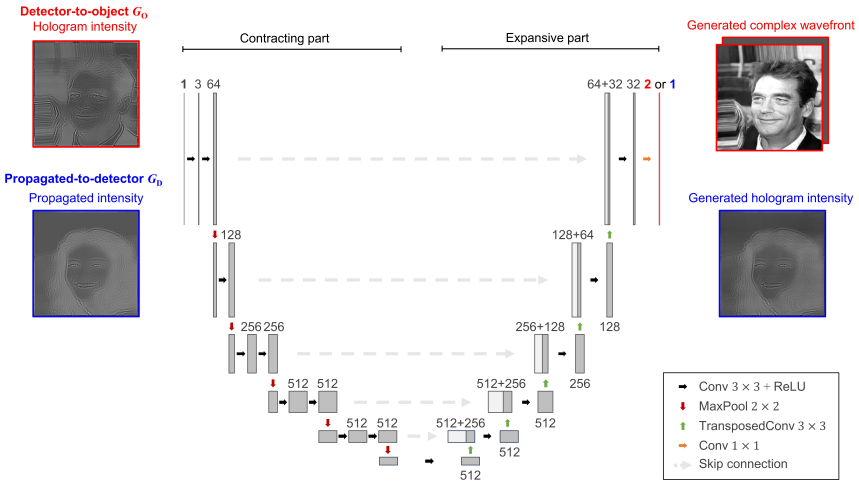

Fig. S1. Generator architecture of PhaseGAN. The contracting and expansive path of the two PhaseGAN generators ($\boldsymbol{G}_\mathrm{O}$ and $\boldsymbol{G}_\mathrm{D}$) are depicted here. The phase-retrieval generator $\boldsymbol{G}_\mathrm{O}$ maps the hologram intensity in the detector domain to a complex wavefront in the object domain, generating a two-channel output from a single channel input. $\boldsymbol{G}_\mathrm{D}$ maps the propagated intensity in the detector plane to the measured detector intensity, i. e., it maps one input channel to one output channel.

Nvidia Tesla V100 SXM2 GPU with 32 GB of RAM for all the training involved in this work. For a given dataset, the speed of training is dependent on various elements, including the network architecture, data loader, mini-batch size, and the memory of the devices. For the training of metallic foam dataset using 32 GB memory and mini-batch size of 40, it took less than 10 hours to go through 100 epochs. The reconstruction process is less time-consuming. It took 20 ms to reconstruct 50 frames. Each generator contains 22.93 million learnable parameters, while the discriminators have 2.76 M. The model sizes of the well-trained generator and discriminator are 460 MB and 55 MB, respectively.

We provide the PyTorch implementation of PhaseGAN, which is based on the architectures from [4] and [2]. The PhaseGAN implementation is available at GitHub.

2. PhaseGAN results summary

This section presents the training strategy and results obtained for the validation (synthetic) and time-resolved metallic-foam experiments.

PhaseGAN is an unpaired phase-reconstruction approach. To train on unpaired datasets, PhaseGAN needs two cycles that use either detector measurements or phase-reconstructed objects as input. Each of these cycles is required to be consistent, i. e., the input should be recovered at the end of the cycle. The two PhaseGAN cycles with their intermediate steps for the validation and experimental datasets are shown in Fig. S2. These steps include two generators $\boldsymbol{G}_\mathrm{O}$ and $\boldsymbol{G}_\mathrm{D}$. The $\boldsymbol{G}_\mathrm{O}$ learns the mapping between the measured detector intensity (I) to the object complex wavefront (Ψ). The $\boldsymbol{G}_\mathrm{D}$ learns the mapping between the estimated intensity on the detector plane to the actual measured intensity. Another intermediate step includes the physics of the image formation via the propagator ($\boldsymbol{H}$). $\boldsymbol{H}$ propagates the complex wavefront from the object plane to the detector plane. The inclusion of the propagator is crucial to enhance the performance of

the phase-reconstructions obtained by PhaseGAN. Finally, Fig. S2 evidences the capability of PhaseGAN to fulfil the cycle consistency.

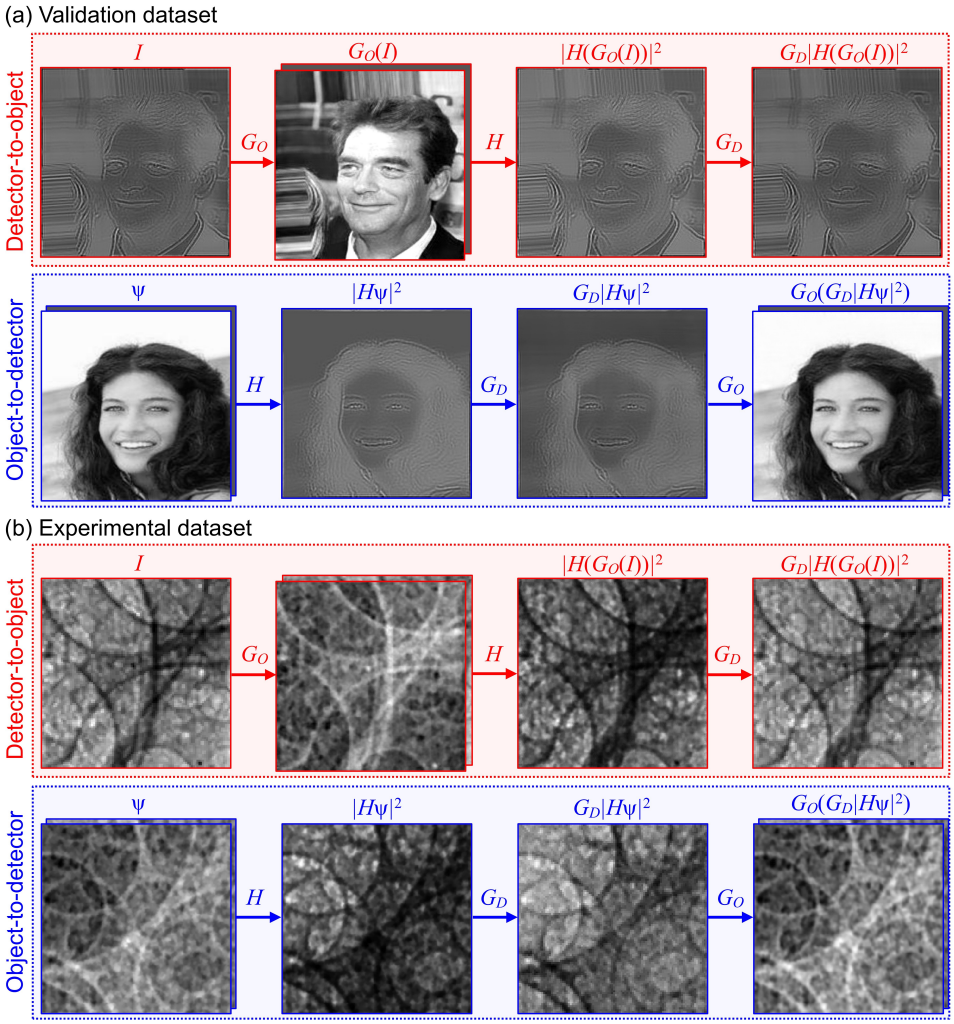

Fig. S2. PhaseGAN cycle-consistency illustration for the (a) validation and (b) experimental datasets. Inside the red box, the cycle from intensity measurements back to intensity measurements is shown. The blue box depicts the complex-wavefront closed cycle. The intermediate steps within each cycle are illustrated. Those intermediate steps use $\boldsymbol{G}_{\mathrm{O}}$, $\boldsymbol{H}$, and $\boldsymbol{G}_{\mathrm{D}}$.

We have performed several tests to understand the capabilities of PhaseGAN compared to state-of-the-art DL approaches. Specifically, we have compared PhaseGAN to: i) classical supervised learning approach using `paired` datasets, ii) adversarial supervised learning with paired datasets using a `pix2pix` method [5], iii) standard `CycleGAN` [4], and iv) `PhaseGAN` with $\alpha_{\mathrm{FRC}} = 0$, denoted by `PhaseGAN`*. For more details about the used methods, the reader is referred to the main text. All these approaches use the same $\boldsymbol{G}_{\mathrm{O}}$ architecture to retrieve the phase.

One of the most simple tests to understand its capabilities is to look at phase profiles over areas difficult to reconstruct, i. e., regions with a high variation of the phase profile over a small area. The results for three line profiles are shown in Fig. S3. It can be seen that all five methods are

capable of reconstructing the homogeneous regions seen in the reference or oracle wavefield. However, the main discrepancies were observed around the object edges.

Second, we report the statistical distributions of three quality metrics L_2 norm, DSSIM, and FRCM for the five DL approaches. For more details about these metrics, the reader is referred to the main text. Smaller values of these three metrics correspond to better reconstructions. Conversely, larger values evidence worse reconstructions. The distributions over 1000 validation images for the L_2 norm, DSSIM, and FRCM, are shown in Fig. S4(a), (b), and (c), respectively. For each metric, we also include the best-performed and the worst-performed validation images of each DL method. The left side of the figure depicts the ranked distribution for each metric from smaller to larger values. The ranked distributions are independent for each of the DL methods, e. g., the smallest value for a given metric and method does not have to be obtained from the same input image as for another method with the same metric. The image patches on the left (right) side of each ranked distribution show the best (worst) phase-retrieved results for each DL approach. The frame colour follows the legend colour code for each method. As expected, most of the methods perform better for clean-background, low-contrast, and less-detail images than images with messy backgrounds, high contrast, and full of details. On the right side of Fig. S4, the kernel-density estimations are depicted for each of the methods and metrics. These distributions are calculated over the logarithmic distribution of values to enhance the differences between the methods. One can see that `PhaseGAN` outperforms `CycleGAN`* and performs at the level of current-state-of-the-art paired DL approaches when applied to the phase problem. Furthermore, by comparing the phase profile and error distribution plots between `PhaseGAN` and `PhaseGAN`*, it becomes clear that the inclusion of FRC loss in the training process helped `PhaseGAN` to achieve better results.

Finally, we display the PhaseGAN reconstruction results of five selected frames extracted from a time-resolved X-ray imaging experiment in Fig. S5. This experiment studied the coalescence of metallic-foam bubbles. This is a crucial process that determines the final structure of the metallic foam [6]. The Intensity row corresponds to measurements performed with a MHz X-ray imaging acquisition system based on a Shimadzu HPV-X2 camera. This system was capable of recording single X-ray pulses provided by the Advanced Photon Source (APS). The phase and attenuation rows correspond to the phase-retrieved results from PhaseGAN, which cannot be provided by current methods. The last row in Fig. S5 shows a schematic illustration of the coalescence process.

PhaseGAN provided a satisfactory solution for this condition, which can provide almost real-time (kHz) phase reconstructions avoiding experimental artifacts in the absence of paired image examples. PhaseGAN can also work as an alternative to the traditional iterative phase reconstruction methods in the need for large volumes of data and rapid reconstructions.

3. Experimental validation with standard objects

The experimental validation of unpaired algorithms can be cumbersome in scenarios where traditional phase-reconstruction approaches do not apply. In this section, we present an experimental validation and comparison of the unpaired algorithms using well-known objects. Using soda-lime glass spheres as well-known objects, we can generate ground-truth images from measured holograms without requiring phase reconstructions. The here studied holograms were measured at the European X-ray free-electron laser (EuXFEL) with 9.3 keV X-ray photons [7]. A total of 893 holograms were measured with a propagation distance of 38 cm and an illuminated area of 855×637 pixels, where the effective pixel size was 520 nm. Each of these holograms suffers from the shot-to-shot noise described in the main text as a consequence of the stochastic process that generates the X-ray free-electron laser (XFEL) radiation [8]. Examples of detector frames are depicted in Fig. S6(a,b). This stochastic noise makes it difficult to obtain conventional phase reconstructions. However, one can obtain phase reconstructions by using unpaired approaches

such as PhaseGAN.

To perform unpaired phase reconstructions, we trained CycleGAN, PhaseGAN*, and PhaseGAN using two independent datasets. The first dataset consisted of 3572 patches from the measured holograms with 256 × 256 pixels. To obtain this dataset, we performed data augmentation by randomly flipping, transposing, and rotating those patches. The second dataset was generated by randomly positioning 80 μm soda-lime glass spheres without physical superposition and using their well-known index of refraction at 9.3 keV [9]. The training process with the two consistent cycles for PhaseGAN and PhaseGAN* is depicted in Fig S6(a). We used the same phase-retrieval ($\boldsymbol{G}_{\mathrm{O}}$) and detector ($\boldsymbol{G}_{\mathrm{D}}$) networks as shown in Fig. S1 for all three methods. To train these methods, we used 4 × Nvidia Tesla V100 SXM2 GPU with 32 GB of RAM running in parallel to have a mini-batch size of 150. The learning rates of the generator and the discriminator were set to be 0.0002 and 0.0001, respectively, and were decayed by 10 after every 150 epochs. It took 12 hours to complete the training of 500 epochs for each method. We used α_{Cyc} =10 and α_{FRC} =1 to train PhaseGAN. However, we set α_{FRC} =0 for PhaseGAN* and CycleGAN. To avoid aliasing artifacts due to the physical propagation while training PhaseGAN* and PhaseGAN, we applied a 256 × 256 Tukey window (tapered cosine) with cosine fraction α = 0.25 to the object images and zero-padded them into 512 × 512 before propagation. The effects of this window in the training process are shown in Fig. S6(a).

To validate our results, we generated ground truth images or oracles by estimating the center of the concentric holographic patterns. Then, we estimated the complex wave projection of soda-lime glass spheres with 80 μm diameter when imaged with 9.3 keV X-ray photons. The process to generate ground truth images is depicted in Fig. S6(b). A total of 100 validation images were used for the results presented here. The reconstructed object images were cropped into 200 × 200 pixels for validation to avoid edge artifacts due to the physical propagation. Figure S6(c) presents the line profile over a spherical feature for the three different methods. Finally, we evaluate the L_2, the DSSIM, and the FRCM metrics for all three methods. The kernel density distributions for the three methods and three metrics are presented in Fig. S6(d). The mean and standard deviation of these figures of merit are reported in Table S1. One can observe that PhaseGAN clearly outperforms the other two methods. Thus, we conclude that adding the FRC loss and the physics of the image formation is crucial to increase the performance of unpaired and unsupervised phase reconstructions. Furthermore, these results are consistent with our synthetic validation.

Table S1. Experimental validation with standard objects. The mean and sigma of the metric distributions for the three unpaired methods are reported.

	L_2		DSSIM		FRCM	
	$\times 10^{-2}$		$\times 10^{-2}$		$\times 10^{-2}$	
	Mean	Std	Mean	Std	Mean	Std
PhaseGAN	1.16	0.56	2.47	0.68	2.02	1.37
PhaseGAN*	1.76	0.81	3.24	0.75	2.67	1.47
CycleGAN	3.23	4.88	3.57	2.17	4.71	4.89

References

1. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention,* (Springer, 2015), pp. 234–241.

2. V. Iglovikov and A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation," arXiv preprint arXiv:1801.05746 (2018).
3. 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.
4. 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.
5. 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.
6. F. García-Moreno, M. Mukherjee, C. Jiménez, A. Rack, and J. Banhart, "Metal foaming investigated by x-ray radioscopy," Metals **2**, 10–21 (2012).
7. C. Pellegrini, A. Marinelli, and S. Reiche, "The physics of x-ray free-electron lasers," Rev. Mod. Phys. **88**, 015006 (2016).
8. 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**, 70–73 (1994).
9. B. Henke, E. Gullikson, and J. Davis, "X-ray interactions: Photoabsorption, scattering, transmission, and reflection at e = 50-30,000 ev, z = 1-92," At. Data Nucl. Data Tables **54**, 181–342 (1993).

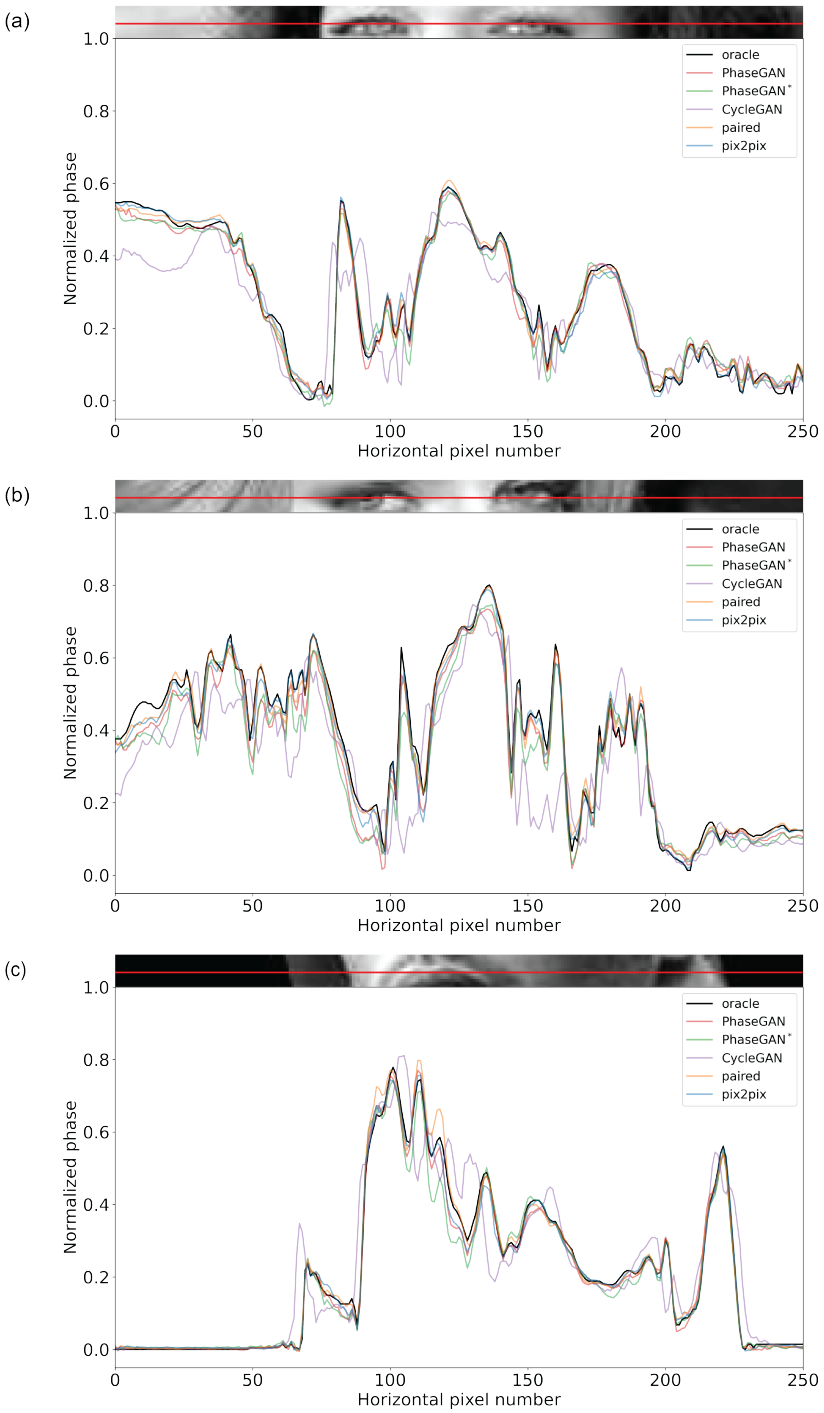

Fig. S3. Phase-reconstructed line profiles by the five DL methods for three independent validation samples. Segments of the oracle images are shown on top of each line-profile plots, where the red path indicates the depicted line profile.

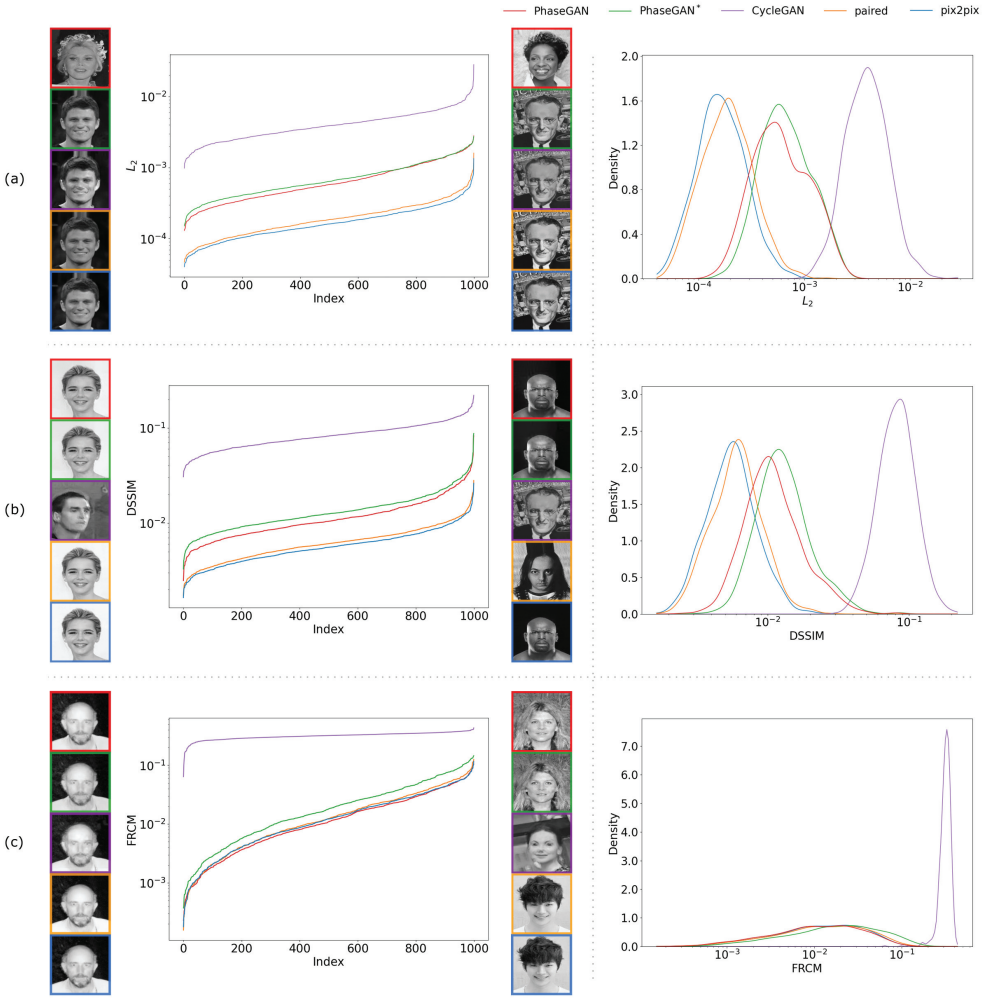

Fig. S4. Comparison of the ranked distribution (left) and the kernel density estimation (right) of PhaseGAN (red), PhaseGAN* (green), CycleGAN (purple), paired (orange), and pix2pix (blue) according to L_2 norm (a), DSSIM (b), and FRCM (c).Insets shows the best and worst reconstructed images for each DL method corresponding to the three different metrics. The color of the insets box indicates the DL method being used.

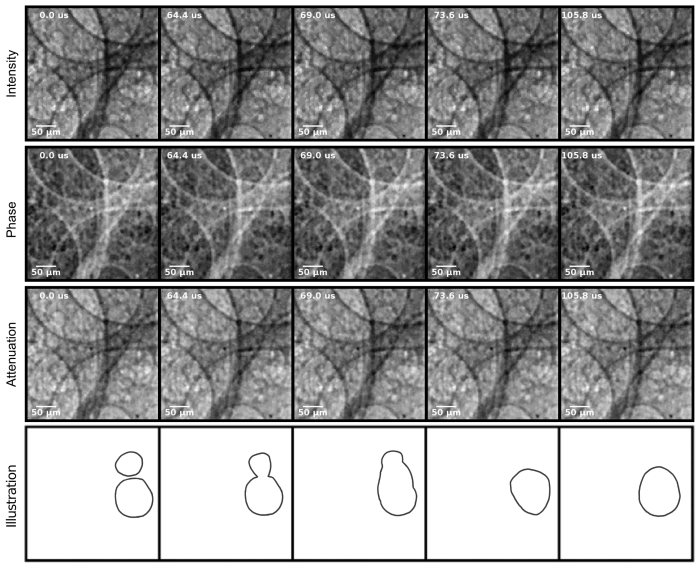

Fig. S5. Clips of the supplementary movie (see Visualization 1, 2, and 3) showing two bubble coalescence of metallic foam. The sketches on the bottom row illustrate this process.

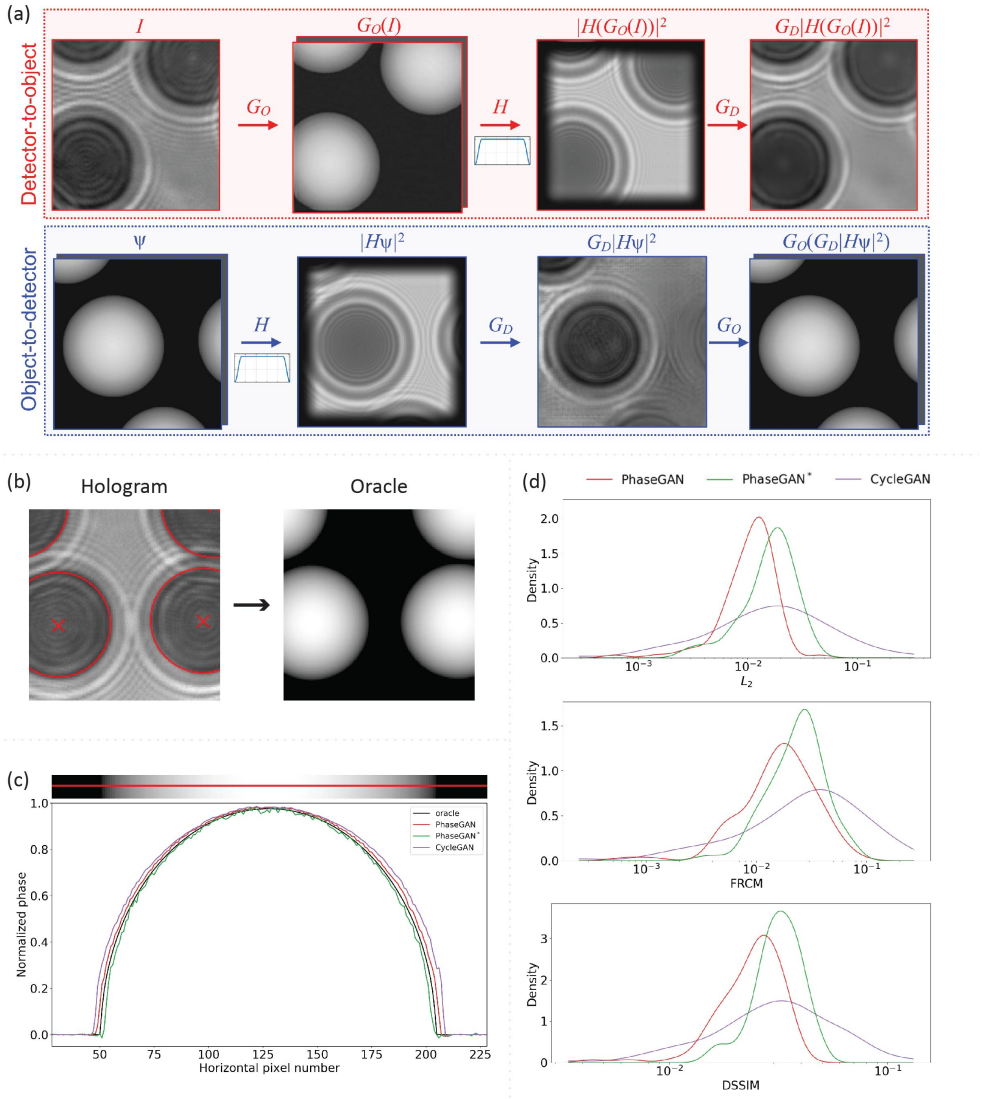

Fig. S6. Sphere validation. (a) PhaseGAN cycle consistency for the experimental validation dataset. (b) Oracle or ground truth generation for a given hologram. (c) Phase-reconstructed line profiles for the three unpaired DL methods. (d) Kernel density estimation for the L_2 norm, DSSIM, and FRCM metrics for PhaseGAN (red), PhaseGAN* (green), and CycleGAN (purple).