

DiffPR: Diffusion-Based Phase Reconstruction via Frequency-Decoupled Learning

Yi Zhang¹

Abstract Oversmoothing remains a persistent problem when applying deep learning to off-axis quantitative phase imaging (QPI): end-to-end U-Nets favour low-frequency content and under-represent fine, diagnostic detail. We trace this issue to *spectral bias* and show that the bias is reinforced by the high-level skip connections that feed high-frequency features directly into the decoder. Removing those deepest skips—thus supervising the network only at a low resolution—significantly improves generalisation and fidelity. Building on this insight, we introduce **DiffPR**, a two-stage frequency-decoupled framework. *Stage 1*: an *asymmetric* U-Net with cancelled high-frequency skips predicts a 1/4-scale phase map from the interferogram, capturing reliable low-frequency structure while avoiding spectral bias. *Stage 2*: the upsampled prediction, lightly perturbed with Gaussian noise, is refined by an unconditional diffusion model that iteratively recovers the missing high-frequency residuals through reverse denoising. Experiments on four QPI datasets (*B-Cell*, *WBC*, *HeLa*, *3T3*) show that DiffPR outperforms strong U-Net baselines, boosting PSNR by up to 1.1 dB and reducing MAE by 11%, while delivering markedly sharper membrane ridges and speckle patterns. The results demonstrate that cancelling high-level skips and delegating detail synthesis to a diffusion prior is an effective remedy for the spectral bias that limits conventional phase-retrieval networks.

1. Introduction

2. Introduction

Off-axis quantitative phase imaging (QPI) provides high-sensitivity, label-free measurement of optical phase delay and has shown strong potential in applications such as live cell imaging (Park et al., 2006; Pandey et al., 2019), morphology analysis (Ryu et al., 2021; Popescu et al., 2006), and cellular mechanism studies (Pandey et al., 2019; Popescu et al., 2006). However, conventional phase reconstruction in off-axis QPI requires computationally expensive post-processing and background calibration, limiting its real-time deployment (Bhaduri et al., 2014; Park et al., 2018;

¹Institute of Data Science, University of Hong Kong. Correspondence to: Yi Zhang <yizhang101@connect.hku.hk>.

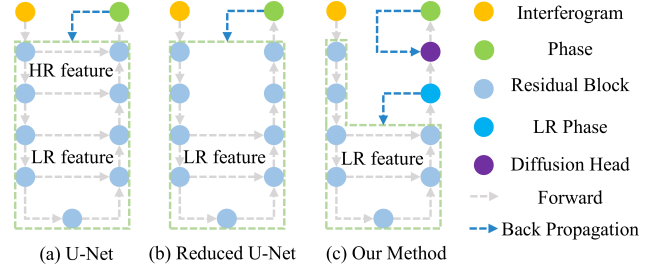


Figure 1. Network comparison. (a) Vanilla U-Net fuses both low- and high-resolution (HR) features via full skip connections; (b) Reduced U-Net removes top-level skips, forcing the decoder to rely on low-resolution (LR) information only; (c) Our DiffPR replaces the HR skip path with a diffusion head that synthesises high-frequency details from the predicted LR phase. Dashed blue arrows indicate gradients flowing from HR supervision.

Niu et al., 2020). To address this, recent efforts have applied deep learning to directly map interferograms to phase maps (Wang et al., 2019b; Chang et al., 2020; Lai et al., 2021; Shi et al., 2019; Yao et al., 2020), enabling faster, calibration-free reconstruction.

Despite these advances, deep neural networks often struggle to recover fine structural details. A well-documented phenomenon in modern deep learning is the spectral bias: neural networks tend to learn low-frequency components significantly faster than high-frequency ones during training (Rahaman et al., 2019). This mismatch becomes particularly problematic for high-resolution phase reconstruction, where high-frequency phase details are essential but difficult to learn directly.

In this work, we investigate how network design impacts frequency learning in deep learning-based phase reconstruction. Specifically, we revisit the common use of U-Net architectures (Ronneberger et al., 2015; Rivenson et al., 2018; Zhang et al., 2021), which fuse multi-level features through skip connections. While lower-level skip connections primarily carry low-frequency content, higher-level skip connections often reintroduce high-frequency information. Surprisingly, we find that removing these high-level skip connections (see Fig. 1 b), which suppresses direct high-frequency supervision, actually improves reconstruction.

tion fidelity and numerical performance. This suggests that simultaneously learning low- and high-frequency content, via architectural design, can be detrimental in this task.

Motivated by this observation, we propose DiffPR, a novel frequency-decoupled framework for phase reconstruction. DiffPR first uses a feedforward neural network to predict a low-resolution phase map from the interferogram. This low-resolution output is then upsampled to the target resolution and perturbed with mild Gaussian noise. Finally, an unconditional diffusion model transforms this noisy input into a high-resolution phase map through a reverse denoising process. Notably, diffusion models are inherently well-suited for learning high-frequency details. During training, they are explicitly tasked with recovering the original signal from inputs corrupted by small amounts of noise—effectively enforcing a learning bias toward fine-grained structures. This makes them particularly powerful for restoring the high-frequency components that standard networks struggle to learn. By decoupling frequency learning and leveraging this bias, our approach achieves sharper, more accurate reconstructions without the need for end-to-end high-frequency supervision.

Our contributions are summarized as follows:

1. We empirically demonstrate that removing high-frequency skip connections in U-Net leads to improved phase reconstruction, revealing that direct joint learning of low- and high-frequency features can hinder performance.
2. We introduce DiffPR, a diffusion-based framework that decouples frequency learning: a feedforward model reconstructs low-frequency phase content, and a diffusion model generates the high-frequency residuals, yielding high-fidelity phase maps with sharper structures.

3. Related Work

3.1. Phase Retrieval on Off-Axis Phase Retrieval

Off-axis quantitative phase imaging (QPI) has matured from its optical foundations into a versatile tool for label-free bio-imaging. Early diffraction-phase microscopy demonstrated sub-nanometer path-length sensitivity for live cells and laid the groundwork for subsequent biomedical applications (Park et al., 2006; Popescu et al., 2006; Bhaduri et al., 2014). Follow-up studies broadened QPI’s reach to morphology analysis, mechano-molecular phenotyping and portable systems, enabling in-field metrology and point-of-care diagnostics (Park et al., 2018; Niu et al., 2020; Pandey et al., 2019; Ryu et al., 2021). Together, these efforts established the need for rapid, high-quality phase reconstruction pipelines to unlock QPI’s real-time potential.

To overcome the computational burden of Fourier-based or iterative solvers, the community has increasingly embraced deep learning. Rivenson *et al.* first showed that a CNN could directly map interferograms to phase maps, suppressing twin-image artefacts with millisecond inference times (Rivenson et al., 2018). Subsequent U-Net variants achieved real-time digital focusing and phase compensation in off-axis microscopy (Zhang et al., 2018; Wang et al., 2018), while one-step networks tackled the notoriously difficult phase-unwrapping task under heavy noise (Wang et al., 2019a;b). Extensions such as PhaseGAN introduced adversarial loss to mitigate the characteristic over-smoothing of ℓ_1 -trained models (Zhang et al., 2021), whereas data-driven aberration modelling enabled truly calibration-free QPI (Chang et al., 2020). Despite these advances, conventional encoder–decoder designs still struggle to recover high-frequency details, motivating our frequency-decoupled DiffPR framework, which predicts reliable low-frequency structure first and then synthesises high-frequency residuals with a generative prior.

3.2. Diffusion Models and High-Frequency Detail Restoration

Denoising Diffusion Probabilistic Models (DDPM) first framed generation as a gradual denoising process, showing that repeatedly predicting clean data from slightly noised inputs can synthesize sharp images (Ho et al., 2020b). Subsequent work demonstrated that diffusion models not only rival but surpass GANs in perceptual fidelity, especially on edge-rich datasets (Dhariwal & Nichol, 2021). Nichol and Dhariwal further improved the noise schedule and sampling procedure, boosting detail preservation in generated outputs (Nichol & Dhariwal, 2021). Score-based generative modeling reformulated DDPMs as stochastic differential equations, providing theoretical insight into how successive denoising naturally reconstructs fine spatial frequencies (Song et al., 2020). Latent Diffusion Models compress pixel space yet still reproduce high-frequency textures at megapixel resolution, underscoring the framework’s efficiency in retaining details after heavy dimensionality reduction (Rombach et al., 2022).

Targeted restoration tasks highlight the bias toward high-frequency learning: SR3 recovers $\times 8$ super-resolution facial details via iterative diffusion refinement (Saharia et al., 2022b), while *Palette* generalises the paradigm to diverse image-to-image translations with texture-faithful outputs (Saharia et al., 2022a). Variational Diffusion Models show that carefully chosen noise schedules and variational objectives yield state-of-the-art likelihoods without sacrificing edge sharpness (Kingma et al., 2021), and EDM systematically analyses how sampler preconditioning accentuates high-frequency convergence (Karras et al., 2022).

Algorithm 1 Inference pipeline of DiffPR

Require: interferogram I ; predictor f_θ ; diffusion model ϵ_θ ; noise schedule $\{\alpha_t\}_{t=1}^T$; corruption level t_{corrupt} ; Reverse Diffusion Solver \mathcal{R}

- 1: **Low-frequency prediction:** $\hat{\phi}^{\text{LR}} \leftarrow f_\theta(I)$
- 2: **Initialisation:** $\tilde{\phi}_T \leftarrow \text{Up}(\hat{\phi}^{\text{LR}}) + \sigma_{t_{\text{corrupt}}} \epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$
- 3: **for** $t = t_{\text{corrupt}}, \dots, 0$ **do**
- 4: **Predictor step:** $\tilde{\phi}_{t-1} \leftarrow \mathcal{R}(\tilde{\phi}_t, t, \epsilon_\theta)$
- 5: **end for**
- 6: **return** $\hat{\phi}^{\text{HR}} \leftarrow \tilde{\phi}_0$

3.3. Spectral Bias of Deep Neural Networks

A growing body of work shows that deep networks possess a *spectral bias*: during training they fit low-frequency components far earlier than high-frequency ones. Rahaman *et al.* first quantified this effect with Fourier analysis, demonstrating that ReLU networks converge in a frequency-dependent manner and struggle to represent rapid oscillations without extensive parameter tuning (Rahaman *et al.*, 2019). Xu *et al.* verified the phenomenon on high-dimensional benchmarks and provided a theoretical explanation based on activation-function regularity (Xu *et al.*, 2019). To mitigate the bias, Tancik *et al.* introduced Fourier feature positional encodings that pre-condition the neural tangent kernel and enable multi-layer perceptrons to capture high-frequency signals (Tancik *et al.*, 2020).

Recent analyses interpret denoising diffusion probabilistic models (DDPMs) as *approximate autoregressive generators in the frequency domain*: the reverse process first reconstructs low-frequency coefficients and then sequentially refines higher-frequency bins (Dieleman, 2024; Russo, 2025). This spectral progression biases diffusion priors toward faithful high-frequency restoration, offering a natural complement to our DiffPR framework, which delegates high-frequency synthesis to an unconditional diffusion stage while maintaining reliable low-frequency structure through a feed-forward predictor.

4. Method

Our DiffPR framework consists of two stages: **(i) Low-resolution phase prediction**, where an asymmetric U-Net maps the input interferogram to a $\frac{1}{4}$ -scale phase map; **(ii) High-frequency refinement**, where a symmetric diffusion U-Net generates the final high-resolution phase by reversing a stochastic denoising process. We denote an interferogram by $I \in \mathbb{R}^{H \times W}$, the ground-truth phase by ϕ^{HR} , and the $\frac{1}{4}$ -scale phase by $\phi^{\text{LR}} = \text{Down}(\phi^{\text{HR}})$.

4.1. Low-Resolution Phase Predictor

Network architecture. We employ an *asymmetric* U-Net f_θ whose encoder has four down-sampling blocks (stride 2 convolutions) and whose decoder has only two up-sampling blocks. Consequently, the network output $\hat{\phi}^{\text{LR}} = f_\theta(I)$ has spatial dimensions $\frac{1}{4}H \times \frac{1}{4}W$.

Training loss. The predictor is trained with a mean-squared error (MSE) objective:

$$\mathcal{L}_{\text{LR}}(\theta) = \mathbb{E}_{(I, \phi^{\text{HR}})} \left[\|f_\theta(I) - \phi^{\text{LR}}\|_2^2 \right]. \quad (1)$$

4.2. Diffusion Prior for High-Frequency Restoration

Network architecture. The diffusion prior is a *symmetric* U-Net ϵ_ω identical in depth on the encoder and decoder sides, operating at the full resolution $H \times W$.

Forward diffusion process. Diffusion models first define a forward diffusion process to perturb the data distribution p_{data} to a Gaussian distribution. Formally, the diffusion process follows an SDE $d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t) + g(t)d\mathbf{w}$, where $d\mathbf{w}$ is the Brownian motion and t flows forward from 0 to T . The calculated analytic solution of this diffusion process gives a transition distribution $p_t(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t | \alpha_t \mathbf{x}_0, \sigma_t^2 \mathbf{I})$, where $\alpha_t = e^{\int_0^t f(s)ds}$ and $\sigma_t^2 = 1 - e^{-\int_0^t g(s)^2 ds}$. In the typical variance-preserving diffusion schedule, \mathbf{f} and g are designed such that $\lim_{t \rightarrow 0} p_t(\mathbf{x}) = p_{\text{data}}(\mathbf{x})$ and $\lim_{t \rightarrow T} p_t(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \mathbf{0}, \mathbf{I})$.

Denoising score matching loss. Diffusion models sample data by reversing this diffusion process, where $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is required. To learn this term, a neural network s_θ is trained to minimize an empirical risk by marginalizing $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t | \mathbf{x}_0)$, leading to the following loss:

$$L(\theta) = \mathbb{E}_{t \sim U(0,1), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \sum_{n=1}^N \|s_\theta(\alpha_t \mathbf{x}_n + \sigma_t \epsilon, t) + \epsilon / \sigma_t\|^2.$$

To further balance the diffusion loss at different t 's, people usually adopt loss reweighing (Karras *et al.*, 2022) or an alternate objective using ϵ -prediction (Ho *et al.*, 2020a; Nichol & Dhariwal, 2021), leading to the following well-known denoising score matching (DSM) loss:

$$L(\theta, t) = \mathbb{E}_{\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \sum_{n=1}^N \|s_\theta(\alpha_t \mathbf{x}_n + \sigma_t \epsilon, t) - \epsilon\|^2.$$

where $s_\theta(\cdot, t)$ can be viewed as the learned score function at time t .

Reverse SDE sampling. Given an estimated low-resolution phase map, we first corrupt it by the forward diffusion process to certain timestep t . $\tilde{\phi}_T = \alpha_t \text{Up}(\hat{\phi}^{\text{LR}}) +$

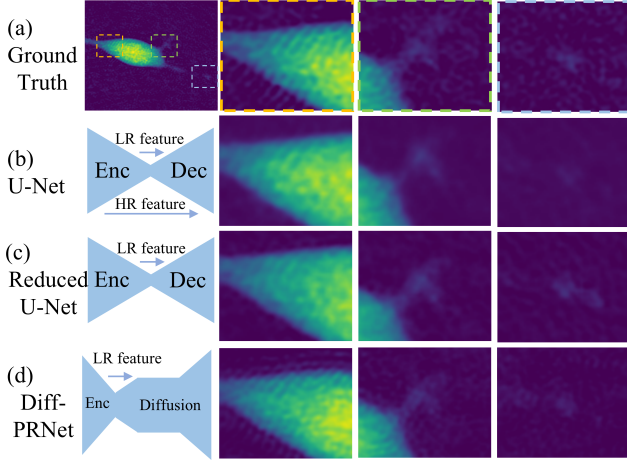


Figure 2. **Qualitative comparison.** (a) Ground-truth phase and zoom-in regions. (b) U-Net results are overly smooth; (c) Reduced U-Net restores better contrast but misses texture; (d) DiffPR reconstructs sharp, noise-free details.

$\sigma_t \epsilon$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$, Starting from this noise-perturbed phase map, we sample from the diffusion model by applying a reverse-time SDE which reverses the diffusion process (Anderson, 1982):

$$d\mathbf{x}_t = [\mathbf{f}(\mathbf{x}_t) - g(t)^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)]dt + d\bar{\mathbf{w}},$$

where $d\bar{\mathbf{w}}$ is the Brownian motion and t flows forward from T to 0.

5. Experiments

Table 1. Numerical Criteria for the Networks on Different Samples. The best results are in bold; \times indicates that the model failed to converge and predicted a blank background.

	Metric	B Cell	WBC	HeLa	3t3
U-Net	PSNR	\times	\times	28.59	34.41
	SSIM	\times	\times	0.6119	0.7687
	MAE	\times	\times	0.0317	0.0138
Reduced U-Net	PSNR	46.64	45.70	28.88	34.76
	SSIM	0.9506	0.9145	0.7080	0.8278
	MAE	0.0028	0.0035	0.0280	0.0119
Diff-PRNet	PSNR	47.59	46.81	29.48	36.96
	SSIM	0.9659	0.9354	0.7393	0.8443
	MAE	0.0025	0.0028	0.0249	0.0108

We evaluate **DiffPR** on four representative off-axis QPI datasets—*B-Cell*, *WBC*, *HeLa*, and *3T3*—and compare it

against the canonical **U-Net** and our **Reduced U-Net**¹. Results demonstrate that (1) removing high-frequency skip connections already improves fidelity, and (2) adding the diffusion prior further boosts both numerical and visual quality.

5.1. Datasets and Metrics

We construct four off-axis QPI datasets containing interferograms of *B Lymphocytes* (385 samples), *white blood cells* (312), *HeLa* (284), and *3T3* cells (298). Each interferogram is of size 1024×1024 pixels and is normalised to zero mean and unit variance. Target phase maps are linearly rescaled to the $[0, 1]$ range.

Following common practice, we randomly split each dataset into 80 % training, 10 % validation, and 10 % test sets. Performance is reported on the test split with **PSNR** (dB), **SSIM**, and **MAE** (lower is better).

5.2. Implementation Details

Network variants. We train three models on all datasets: (i) the canonical **U-Net**, (ii) the proposed **Reduced U-Net**, and (iii) our full **DiffPR** (which uses the Reduced U-Net as low-resolution predictor and a diffusion prior).

Training protocol for U-Net variants. Input–target pairs are randomly cropped to 512×512 to accelerate training and reduce GPU memory. All models are optimised with Adam ($\beta_1=0.9$, $\beta_2=0.999$) at a learning rate of 1×10^{-4} and a batch size of 16. For the vanilla U-Net we employ early stopping on the validation PSNR with a patience of 100 epochs.

Diffusion prior. The diffusion model adopts a variance-preserving schedule (Ho et al., 2020a) with $T=1000$ timesteps and is trained for 800k updates. At inference we integrate the reverse SDE with a 100-step predictor–corrector sampler (Alg. 1).

6. Quantitative Results

Table 1 presents PSNR, SSIM, and MAE across all four datasets. On *B-Cell* and *WBC* the baseline U-Net collapses to an all-zero prediction—marked by “ \times ” in the table—because both datasets contain large zero-phase backgrounds. The network minimises loss by staying in this local minimum, and standard gradient descent fails to escape. The experiments reveal several clear patterns:

Removing high-frequency skips helps convergence and fidelity. Eliminating the deepest skip connection forces

¹U-Net without the deepest skip connections; see Fig. 1 (b).

the decoder to rely on low-frequency structure rather than a spurious high-frequency shortcut. As a result, the *Reduced U-Net* converges on every dataset and consistently outperforms the vanilla model. For example, on *B-Cell* PSNR rises from failure to 46.6 dB, and MAE decreases by 13 %.

A diffusion prior further closes the detail gap. Adding the diffusion stage boosts fidelity beyond the *Reduced U-Net*, especially on texture-rich or noisy samples. On *WBC* we observe a +1.1 dB PSNR gain (from 45.7 dB to 46.8 dB) and a corresponding SSIM improvement ($0.914 \rightarrow 0.935$). On *B-Cell* MAE drops from 0.0028 to 0.0025. The iterative denoising process effectively reconstructs membrane granularity that deterministic decoders miss.

Overall, DiffPR delivers the highest PSNR on every dataset—up to +1.1 dB beyond the *Reduced U-Net*—and reduces MAE by as much as 11 %. These results confirm that (i) suppressing direct high-frequency supervision avoids background-dominated minima, and (ii) coupling a low-frequency predictor with a diffusion prior is superior to conventional end-to-end optimisation for high-fidelity phase reconstruction.

6.1. Qualitative Results

Figure 2 illustrates visual differences on representative crops. Grey dashed boxes in the ground truth highlight three challenging regions: fine ridges, a low-contrast cytoplasmic boundary, and speckle-like noise.

Vanilla U-Net (Fig. 2b). Although edges are roughly aligned, high-frequency texture is missing; ridge amplitude is attenuated and intracellular speckles collapse into smooth blobs—classic symptoms of spectral bias.

Reduced U-Net (Fig. 2c). Removing top-level skips sharpens global shape and attenuates ringing, but residual blur remains around membrane folds and the faint cytoplasmic boundary is still over-smoothed. This suggests that frequency decoupling helps preserve macroscopic structure yet cannot fully regenerate lost fine features.

DiffPR (Fig. 2d). Our method faithfully reconstructs the sinusoidal ridges, restores the weak boundary with correct phase gradient, and reproduces speckle patterning without hallucinated noise. Because the diffusion prior progressively denoises from a low-frequency initialisation, it focuses its capacity on residual high-frequency modes rather than re-learning the full signal. The result is a phase map visually indistinguishable from the ground truth while remaining free of artefacts.

These qualitative observations mirror the quantitative improvements in Table 1, reinforcing that DiffPR achieves superior high-frequency fidelity without sacrificing low-frequency accuracy.

7. Conclusion

We have introduced **DiffPR**, a frequency-decoupled framework for off-axis quantitative phase imaging that couples a low-resolution predictor with an unconditional diffusion prior. Our analysis revealed that the ubiquitous *U-Net* architecture, when endowed with deep skip connections, can overfit large background regions and stall in a low-frequency local minimum. Removing the deepest skips (*Reduced U-Net*) already alleviates this issue, confirming that direct supervision of high-frequency features can be harmful. Building on this observation, DiffPR synthesises the missing high-frequency residuals through a reverse SDE, explicitly exploiting the diffusion model’s bias toward detail recovery.

Extensive experiments on four cell datasets demonstrate that DiffPR surpasses both vanilla and *Reduced U-Net* baselines, improving PSNR by up to 1.1 dB and lowering MAE by 11 %. Qualitative comparisons further show that DiffPR reconstructs sub-cellular ridges and speckle patterns with minimal artefacts.

Limitations and future work. DiffPR currently relies on a 100-step sampler, incurring longer inference time than a single forward pass. Future work will explore consistency distillation (Luhman & Luhman, 2021; Liu et al., 2023a; Song et al., 2023; Yin et al., 2024; Zheng et al., 2023; Liu et al., 2023b; Berthelot et al., 2023), and higher-order ODE solvers (Song et al., 2021; Lu et al., 2022a;b) to accelerate sampling. Moreover, extending the framework to volumetric phase tomography and to other ill-posed inverse problems (e.g. fluorescence lifetime imaging) could broaden its impact. Finally, incorporating physics-informed constraints into the diffusion prior is a promising direction for further improving accuracy and interpretability.

References

- Anderson, B. D. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.
- Berthelot, D., Autef, A., Lin, J., Yap, D. A., Zhai, S., Hu, S., Zheng, D., Talbott, W., and Gu, E. Tract: Denoising diffusion models with transitive closure time-distillation. *arXiv preprint arXiv:2303.04248*, 2023.
- Bhaduri, B., Edwards, C., Pham, H., Zhou, R., Nguyen, T. H., Goddard, L. L., and Popescu, G. Diffraction phase microscopy: principles and applications in materials and life sciences. *Advances in Optics and Photonics*, 6(1): 57–119, 2014.
- Chang, T., Ryu, D., Jo, Y., Choi, G., Min, H.-S., and Park, Y. Calibration-free quantitative phase imaging using

- data-driven aberration modeling. *Optics Express*, 28(23): 34835–34847, 2020.
- Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 2021. arXiv:2105.05233.
- Dieleman, S. Diffusion is spectral autoregression. Blog post, 2024. URL <https://sander.ai/2024/09/02/spectral-autoregression.html>. Accessed: 2025-06-12.
- Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020a.
- Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*, 2020b.
- Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35: 26565–26577, 2022.
- Kingma, D. P., Salimans, T., Poole, B., and Ho, J. Variational diffusion models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2107.00630.
- Lai, X., Xiao, S., Xu, C., Fan, S., and Wei, K. Aberration-free digital holographic phase imaging using the derivative-based principal component analysis. *Journal of Biomedical Optics*, 26(4):046501, 2021.
- Liu, X., Gong, C., and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023a. URL <https://openreview.net/forum?id=XVjTT1nw5z>.
- Liu, X., Zhang, X., Ma, J., Peng, J., et al. InstafLOW: One step is enough for high-quality diffusion-based text-to-image generation. In *The Twelfth International Conference on Learning Representations*, 2023b.
- Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022a.
- Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022b.
- Luhman, E. and Luhman, T. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021.
- Nichol, A. Q. and Dhariwal, P. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.
- Niu, M., Luo, G., Shu, X., Qu, F., Zhou, S., Ho, Y.-P., Zhao, N., and Zhou, R. Portable quantitative phase microscope for material metrology and biological imaging. *Photonics Research*, 8(7):1253–1259, 2020.
- Pandey, R., Zhou, R., Bordett, R., Hunter, C., Glunde, K., Barman, I., Valdez, T., and Finck, C. Integration of diffraction phase microscopy and raman imaging for label-free morpho-molecular assessment of live cells. *Journal of biophotonics*, 12(4):e201800291, 2019.
- Park, Y., Popescu, G., Badizadegan, K., Dasari, R. R., and Feld, M. S. Diffraction phase and fluorescence microscopy. *Opt. Express*, 14(18):8263–8268, Sep 2006. doi: 10.1364/OE.14.008263. URL <https://opg.optica.org/oe/abstract.cfm?URI=oe-14-18-8263>.
- Park, Y., Depeursinge, C., and Popescu, G. Quantitative phase imaging in biomedicine. *Nature photonics*, 12(10): 578–589, 2018.
- Popescu, G., Ikeda, T., Dasari, R. R., and Feld, M. S. Diffraction phase microscopy for quantifying cell structure and dynamics. *Optics letters*, 31(6):775–777, 2006.
- Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pp. 5301–5310. PMLR, 2019. URL <https://proceedings.mlr.press/v97/rahaman19a.html>.
- Rivenson, Y., Zhang, Y., Günaydin, H., Teng, D., and Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141–17141, 2018.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv:2112.10752.
- Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.
- Russo, G. Diffusion is autoregression in the frequency domain. Blog post, 2025. URL <https://gianluca.ai/>

- [diffusion-is-frequency-autoregression/](#). Accessed: 2025-06-12.
- Ryu, D., Kim, J., Lim, D., Min, H.-S., Yoo, I. Y., Cho, D., and Park, Y. Label-free white blood cell classification using refractive index tomography and deep learning. *BME Frontiers*, 2021, 2021.
- Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D. J., and Norouzi, M. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH Conference Proceedings*, 2022a. arXiv:2111.05826.
- Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. Image super-resolution via iterative refinement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022b. arXiv:2104.07636.
- Shi, J., Zhu, X., Wang, H., Song, L., and Guo, Q. Label enhanced and patch based deep learning for phase retrieval from single frame fringe pattern in fringe projection 3d measurement. *Optics express*, 27(20):28929–28943, 2019.
- Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=StlgIarCHLP>.
- Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020.
- Song, Y., Dhariwal, P., Chen, M., and Sutskever, I. Consistency models, 2023. URL <https://arxiv.org/abs/2303.01469>.
- Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, 2020. URL <https://arxiv.org/abs/2006.10739>.
- Wang, H., Lyu, M., and Situ, G. eholonet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Opt. Express*, 26(18):22603–22614, Sep 2018. doi: 10.1364/OE.26.022603. URL <https://opg.optica.org/oe/abstract.cfm?URI=oe-26-18-22603>.
- Wang, K., Li, Y., Kemao, Q., Di, J., and Zhao, J. One-step robust deep learning phase unwrapping. *Opt. Express*, 27(10):15100–15115, May 2019a. doi: 10.1364/OE.27.015100. URL <https://opg.optica.org/oe/abstract.cfm?URI=oe-27-10-15100>.
- Wang, K., Li, Y., Kemao, Q., Di, J., and Zhao, J. One-step robust deep learning phase unwrapping. *Optics express*, 27(10):15100–15115, 2019b.
- Xu, Z.-Q. J., Zhang, Y., Luo, T., Xiao, Y., and Ma, Z. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019. URL <https://arxiv.org/abs/1901.06523>.
- Yao, Y., Shu, X., and Zhou, R. Deep learning based phase retrieval in quantitative phase microscopy. In *Unconventional Optical Imaging II*, volume 11351, pp. 76–80. SPIE, 2020.
- Yin, T., Gharbi, M., Zhang, R., Shechtman, E., Durand, F., Freeman, W. T., and Park, T. One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6613–6623, 2024.
- Zhang, G., Guan, T., Shen, Z., Wang, X., Hu, T., Wang, D., He, Y., and Xie, N. Fast phase retrieval in off-axis digital holographic microscopy through deep learning. *Opt. Express*, 26(15):19388–19405, Jul 2018. doi: 10.1364/OE.26.019388. URL <https://opg.optica.org/oe/abstract.cfm?URI=oe-26-15-19388>.
- Zhang, Y., Noack, M. A., Vagovic, P., Fezzaa, K., Garcia-Moreno, F., Ritschel, T., and Villanueva-Perez, P. Phasegan: a deep-learning phase-retrieval approach for unpaired datasets. *Optics express*, 29(13):19593–19604, 2021.
- Zheng, H., Nie, W., Vahdat, A., Aizzadenesheli, K., and Anandkumar, A. Fast sampling of diffusion models via operator learning. In *International conference on machine learning*, pp. 42390–42402. PMLR, 2023.