On the Effectiveness of Physics-Informed Fourier Neural Operators for Holographic Phase Retrieval

Mohammad Saim Khan, Stéphane Cuenat, Rémi Kieber, Maxime Jacquot and Raphaël Couturier
Université Marie et Louis Pasteur, CNRS, Institut FEMTO-ST
Besançon, France
Emails: {mohammad_saim.khan, stephane.cuenat, remi.kieber, maxime.jacquot, raphael.couturier}@umlp.fr

***Abstract*—Digital holography enables marker-free, quantitative phase imaging of transparent biological specimens, yet retrieving phase from intensity-only measurements remains fundamentally ill-posed. We investigate the effectiveness of physics-informed Fourier Neural Operators (FNOs) for holographic phase retrieval through PhysSpec-FNO, a self-supervised spectral boosting framework that derives its entire training signal from the Angular Spectrum Method (ASM) propagator—requiring no paired ground-truth data. The architecture employs a two-stage Generator–Refiner pipeline: the Generator captures dominant low-frequency wavefield structure, while the Refiner corrects high-frequency residuals via concatenated input conditioning. We systematically evaluate PhysSpec-FNO against a supervised Hybrid SpecBoost baseline across four domain configurations, using synthetically generated Random Noise and MS-COCO datasets (100,000 images at 512×512 resolution each). In matched-domain settings, PhysSpec-FNO achieves a peak Structural Similarity Index (SSIM) of 0.846—a 33% relative improvement over the supervised baseline (0.636)—while the supervised method demonstrates superior cross-domain robustness. Validation on real lung tissue specimens further confirms the practical advantage of the physics-informed approach (SSIM 0.509 vs. 0.484). These results illuminate a fundamental specialization–generalization trade-off inherent to physics-informed learning: FNOs constrained by physical propagation laws excel when training and deployment domains are aligned, but sacrifice distributional flexibility. Our findings position physics-informed FNOs as an effective solution for dedicated biological microscopy pipelines where ground-truth annotations are unavailable and domain alignment can be assumed.**

***Index Terms*—holographic phase retrieval, Fourier Neural Operator, self-supervised learning, spectral boosting, physics-informed deep learning, digital holography, biomedical imaging**

I. Introduction

Digital Holography (DH) is a powerful imaging modality that captures both the amplitude and phase of a light wave scattered by a specimen. Unlike conventional microscopy, which records only intensity, imaging in digital holographic microscopy (DHM) provides quantitative phase maps that encode morphological and refractive index information about transparent biological samples—such as live cells, tissue sections, and pathological specimens—without the need for fluorescent labeling or staining [1], [2]. This label-free, non-invasive capability has made DHM increasingly important in biomedical applications including cell counting, disease diagnosis, and real-time monitoring of cellular dynamics.

However, practical holographic sensors record only intensity measurements, discarding the phase component of the complex optical wavefield. Recovering this lost phase—the so-called *phase retrieval problem*—is inherently ill-posed, as multiple complex fields can produce identical intensity patterns [3]. Traditional iterative algorithms, such as the Gerchberg–Saxton algorithm and its variants, rely on alternating projections between spatial and Fourier domains but suffer from slow convergence, stagnation at local minima, and sensitivity to noise [4].

Recent advances in deep learning have produced diverse strategies for the phase retrieval problem. Supervised methods [11] train convolutional networks on paired hologram/phase data, while untrained physics-driven techniques such as the deep image prior [12] optimize a randomly initialized network on a single hologram, exploiting implicit regularization rather than a learned prior. Physics-informed neural operators [9] embed governing wave equations into the training objective, and broader surveys of deep learning for computational imaging [13] catalogue further variants. Among these, the GedankenNet framework [5] stands out by training a single Fourier Neural Operator (FNO) [6] entirely without ground-truth data, leveraging the Angular Spectrum Method (ASM) as a differentiable physics propagator that maps a complex wavefield from the sample plane to the sensor plane via the FFT. The off-axis DHM geometry underlying this propagation model—wherein object and reference beams interfere at a prescribed angle to encode the complex wavefield in a single interferogram—is described in detail in [7]. Independently, Xiao et al. [8] proposed SpecBoost, an ensemble framework that sequentially trains multiple FNOs to mitigate the inherent low-frequency bias of these operators, originally designed for solving partial differential equations (PDEs). While SpecBoost improves prediction accuracy on PDE tasks by up to 71%, its potential for holographic reconstruction—and, more critically, for a self-supervised regime where no residual ground truth is available—has not previously been explored.

In this paper, we propose **PhysSpec-FNO**, a framework that brings the SpecBoost architecture into the domain of self-supervised holographic phase retrieval. Our key contributions are:

1) We adapt the SpecBoost ensemble architecture [8]—originally designed for supervised PDE solving—to

the GedankenNet [5] self-supervised holographic reconstruction pipeline, replacing the single FNO backbone with a two-stage Generator–Refiner architecture that captures both low- and high-frequency components of the complex wavefield.
2) We introduce a fully self-supervised training paradigm for the boosted architecture, eliminating the need for ground-truth phase images by enforcing physics consistency through the ASM propagator, combined with a multi-component loss function (spectral MAE, spatial MSE, and total variation regularization).
3) We conduct a systematic experimental comparison between PhysSpec-FNO and a supervised Hybrid variant across four domain configurations and real biological specimens, demonstrating 33% higher SSIM in matched-domain settings while uncovering a fundamental specialization–generalization trade-off with practical implications for biomedical imaging system design.

II. Related Work

A. Fourier Neural Operators

The FNO, introduced by Li et al. [6], learns mappings between function spaces by parameterizing the integral kernel in the Fourier domain. By performing global convolutions via the Fast Fourier Transform (FFT), FNOs capture long-range dependencies that conventional convolutional architectures miss. This property is particularly advantageous for holographic reconstruction, where the relationship between recorded intensities and the underlying complex field is inherently global in the frequency domain. Subsequent work has extended FNOs to various imaging and PDE-solving tasks [9]. However, Xiao et al. [8] identified a critical limitation: FNOs exhibit a distinct low-frequency bias due to their global Fourier filters and mode truncation, leading to larger errors in high-frequency components of the prediction.

B. GedankenNet: Self-Supervised Holographic Reconstruction

GedankenNet [5] pioneered self-supervised holographic phase retrieval without any ground-truth images. The framework employs a single FNO as its backbone and derives supervision entirely from the physics of wave propagation: the network's predicted complex field is numerically propagated back to the sensor plane using the ASM, and the resulting simulated hologram is compared against the measured input. Crucially, GedankenNet trains on synthetically generated random images bearing no resemblance to real-world samples, yet generalizes remarkably to experimental holograms of biological specimens. This paradigm eliminates the need for paired ground-truth data, which is costly or impossible to acquire in practice.

C. SpecBoost: Spectral Boosting for FNOs

To address FNO's low-frequency bias, Xiao et al. [8] proposed SpecBoost, an ensemble learning framework that sequentially trains two FNOs, as illustrated in Fig. 1. The first FNO ($\mathcal{G}_a$) undergoes standard training and captures low-frequency components effectively. A second FNO ($\mathcal{G}_b$) is then trained on the prediction residuals $r = y - \hat{y}$ of the first, using the concatenation of the input and the first FNO's prediction $[x, \hat{y}]$ as input. The final output combines both predictions: $\hat{y} + \hat{r}$. Originally designed and validated for PDE-solving tasks (Navier–Stokes, Darcy flow), SpecBoost achieved up to 71% improvement in prediction accuracy. However, it was designed as a supervised framework requiring ground-truth labels—its application to self-supervised settings such as holographic reconstruction has not been explored prior to this work.

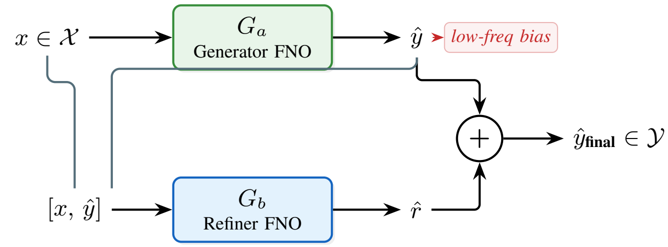

Fig. 1. SpecBoost architecture [8]. The Generator G_a produces an initial estimate $\hat{y}$ exhibiting low-frequency bias. The Refiner G_b receives the concatenated input $[x, \hat{y}]$ and predicts a residual correction $\hat{r}$. The final prediction $\hat{y}_{\text{final}} = \hat{y} + \hat{r}$ captures both low- and high-frequency components.

III. Proposed Method: PhysSpec-FNO

The central idea behind PhysSpec-FNO is to combine the spectral boosting strategy of SpecBoost [8] with the self-supervised holographic paradigm of GedankenNet [5]. GedankenNet employs a *single* FNO2d trained for 2,000 epochs using only a physics-consistency loss. While effective, a single FNO is limited by its inherent low-frequency bias [8], leaving high-frequency phase details under-represented. PhysSpec-FNO addresses this limitation by replacing the single backbone with a two-stage Generator–Refiner pipeline trained entirely through physics-consistency constraints, recovering high-frequency components without requiring ground-truth data.

This transfer is non-trivial: the original SpecBoost [8] trains G_b on the explicit residual $r = y - \hat{y}$, but y is precisely the unknown complex wavefield in our setting, so the residual signal driving its second stage *does not exist*. Our contribution is therefore the training paradigm rather than the architecture: both G_a and G_b are supervised through the same ASM-based physics loop, and a two-phase freeze schedule (Section III-D) replaces the explicit residual signal by structurally constraining G_b to a correction observable only via the propagated hologram.

Fig. 2 summarizes the architectural and training differences between GedankenNet, PhysSpec-FNO, and the supervised Hybrid SpecBoost baseline.

A. Fourier Neural Operator Backbone

Each stage uses a customized FNO2d backbone. The spectral convolution layer performs a 2D FFT, multiplies retained

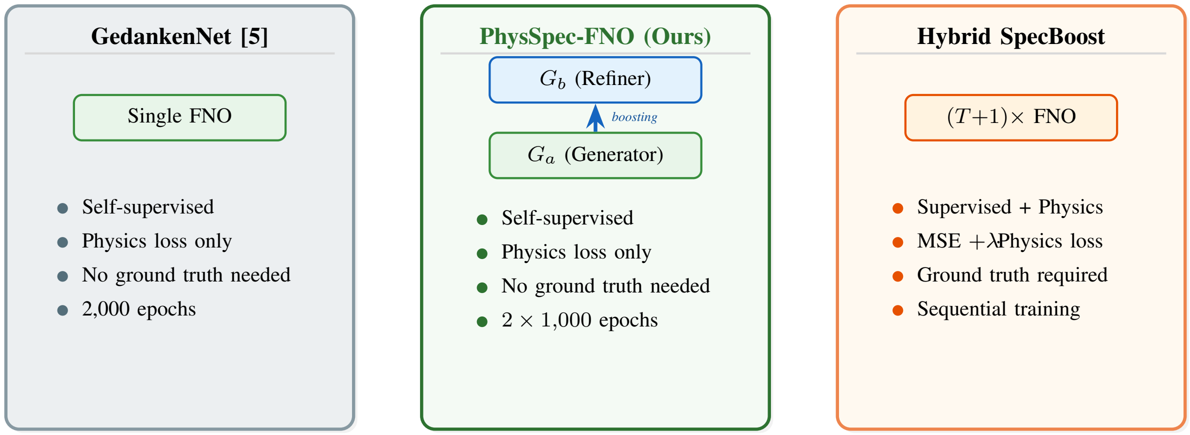

Fig. 2. Comparison of the three frameworks. GedankenNet uses a single FNO with self-supervised physics loss. Our PhysSpec-FNO adds SpecBoost's two-stage spectral boosting while retaining full self-supervision. The Hybrid SpecBoost baseline uses the same architecture with supervised ground-truth training.

Fourier modes ($m_1 = m_2 = 256$) with learnable complex weights R_ϕ, and applies the inverse FFT:

$$\mathcal{K}(v)(x) = \mathcal{F}^{-1}\left(R_\phi \cdot \mathcal{F}(v)\right)(x), \tag{1}$$

where $\mathcal{F}$ denotes the 2D Fourier transform and $\mathcal{F}^{-1}$ its inverse. The backbone consists of 8 multi-scale blocks with scale factors $s \in \{1, 1, 2, 2, 2, 4, 4, 4\}$, each containing 2 sub-layers. Within each block:

$$x^{(l+1)} = \sigma\Big(\mathcal{K}(x^{(l)}) + Wx^{(l)}\Big) + x^{(l)}, \tag{2}$$

where W is a 1×1 convolution and σ is PReLU. Inter-block skip connections and a global long-skip form a three-level residual hierarchy. The input is augmented with 2D positional encoding and projected to $w = 48$ channels; the output produces 2 channels (real and imaginary components).

B. Spectral Boosting Architecture

Following the SpecBoost framework [8], PhysSpec-FNO implements a two-stage spectral boosting pipeline, illustrated in Fig. 3. The first stage, the **Generator** G_a, receives the raw holographic input $x \in \mathbb{R}^{B\times C\times H\times W}$ (where C corresponds to measurements at different axial distances) and produces an initial complex field estimate $\hat{y} = G_a(x)$. Due to FNO's low-frequency bias [8], this estimate primarily captures the dominant low-frequency structure of the complex wavefield.

The second stage, the **Refiner** G_b, is designed to capture the high-frequency residual error that G_a fails to model. Following the SpecBoost design, G_b receives a *concatenated input* consisting of both the original hologram and the Generator's prediction: $[x, \hat{y}]$. This conditioning mechanism provides G_b with information about what G_a has already recovered, allowing it to focus specifically on the overlooked high-frequency details:

$$\hat{y}_{\text{final}} = G_a(x) + G_b([x, \hat{y}])\,. \tag{3}$$

While this architecture mirrors the original SpecBoost [8], the key difference lies in the training paradigm: the original SpecBoost trains both G_a and G_b with supervised ground-truth labels, whereas our PhysSpec-FNO trains using only self-supervised physics-consistency constraints (Section III-C), making it applicable to holographic reconstruction where ground-truth complex fields are unavailable. The total number of learnable parameters in the full PhysSpec-FNO architecture is approximately 54 million.

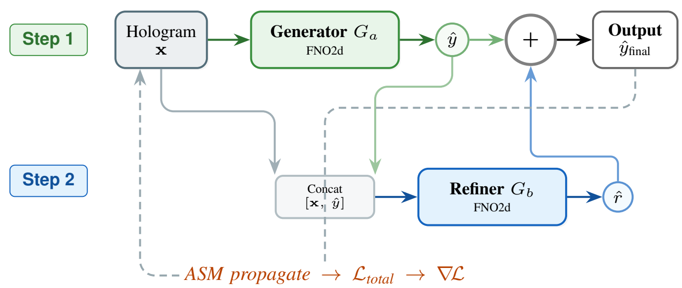

Fig. 3. PhysSpec-FNO architecture. **Step 1**: the Generator G_a maps the input hologram $\mathbf{x}$ to an initial estimate $\hat{y}$. **Step 2**: the Refiner G_b takes $[\mathbf{x}, \hat{y}]$ and predicts a residual $\hat{r}$; the final output is $\hat{y}_{\text{final}} = \hat{y}+\hat{r}$. A physics feedback loop (dashed) propagates $\hat{y}_{\text{final}}$ via the ASM, computes $\mathcal{L}_{\text{total}}$, and back-propagates gradients—no ground truth is needed.

C. Self-Supervised Physics-Consistency Loss

PhysSpec-FNO is trained without paired ground-truth phase images. Instead, supervision is derived entirely from the physics of wave propagation: the predicted complex field is numerically propagated back to the sensor plane and compared against the measured hologram. This self-supervised paradigm is critical for biological microscopy, where ground-truth complex fields are costly or impossible to acquire.

Specifically, given $\hat{y}_{\text{final}}$, a differentiable ASM-based propagator simulates the forward holographic process, computing the propagation of a complex field $U(x, y)$ from plane z_0 to plane z as:

$$U(x, y; z) = \mathcal{F}^{-1}\{\mathcal{F}\{U(x, y; z_0)\} \cdot H(f_x, f_y; \Delta z)\}\,, \tag{4}$$

where f_x, f_y are the spatial frequencies conjugate to x, y, $H(f_x, f_y; \Delta z) = \exp\left(j2\pi\sqrt{k^2 - f_x^2 - f_y^2} \cdot \Delta z\right)$ is the transfer function, $k = n/\lambda$ is the wavenumber ($\lambda = 530$ nm, $n = 1.0$ for air), and Δz is the propagation distance. The simulated holograms are computed at axial distances $z \in \{300, 375\}$ µm with a pixel pitch of 0.3733 µm, considering 10X magnification provided by microscope objective lens in DHM geometry.

The total loss function combines three complementary objectives:

$$\mathcal{L}_{\text{total}} = \alpha\, \mathcal{L}_{\text{MAE}}^{\text{spectral}} + \beta\, \mathcal{L}_{\text{MSE}}^{\text{spatial}} + \gamma\, \mathcal{L}_{\text{TV}}, \tag{5}$$

where the spectral loss operates on the windowed Fourier transform (using a 2D Hann window to suppress spectral leakage), the spatial loss enforces pixel-wise consistency, and the total variation (TV) regularizer promotes piece-wise smooth reconstructions. The weights are set to $\alpha = 0.1$, $\beta = 1.0$, and $\gamma = 20.0$.

D. Two-Phase Training Strategy

To ensure stable convergence of the boosting pipeline, we employ a two-phase training schedule:

- **Phase 1 (Epochs 0–1000):** Only the Generator G_a is trained while the Refiner G_b remains frozen with random initialization. This ensures that G_a first learns a reasonable initial estimate before the Refiner attempts to correct it.
- **Phase 2 (Epochs 1000–2000):** G_a is frozen with its learned weights, and only G_b is trained. During this phase, the Refiner receives gradient signals specifically targeting the errors left by the Generator, enabling it to specialize in residual correction.

Both phases use the Adam optimizer with an initial learning rate of 0.002 and a CosineAnnealingWarmRestarts scheduler with $T_0 = 2$ and multiplicative factor $T_{\text{mult}} = 2$. Gradient clipping with a maximum norm of 1.0 is applied to prevent training instability.

IV. Experimental Setup and Results

A. Dataset and Preprocessing

Following the GedankenNet methodology [5], we use two synthetically generated datasets, each comprising 100,000 samples at 512×512 resolution: (1) **Random Noise Images** (100,000 samples) generated from random Gaussian noise modulated by random phase masks drawn uniformly in $[-\pi, \pi]$, with amplitude clipped to $[0, 1]$; and (2) **COCO Images** (100,000 samples) derived from MS-COCO 2017 [10] grayscale images randomly cropped to 512×512 and linearly mapped to phase values in $[-\pi, \pi]$. For both, the ASM propagator (Eq. 4) simulates holographic recordings at the specified axial distances. The data is split 80%/10%/10% for training/validation/testing.

B. Implementation Details

For reproducibility: both G_a and G_b are FNO2d backbones with $w = 48$ channels, modes $m_1 = m_2 = 256$, 8 multi-scale blocks of scales $\{1, 1, 2, 2, 2, 4, 4, 4\}$ with 2 sub-layers each, PReLU activations, and a three-level residual skip hierarchy (∼54M total parameters). Training uses Adam (lr $= 0.002$, weight decay 10^{-4}, grad-clip norm 1), batch size 2, 250 minibatches/epoch, and CosineAnnealingWarmRestarts ($T_0 = 2$, $T_{\text{mult}} = 2$ for PhysSpec-FNO; $T_0 = 50$ for Hybrid). Loss weights: $\alpha = 0.1$, $\beta = 1.0$, $\gamma = 20.0$, $\lambda = 0.5$. ASM: $\lambda_{\text{light}} = 530$ nm, $n = 1.0$, pixel pitch 0.3733 µm, $z \in \{300, 375\}$ µm.

C. Baseline: Hybrid SpecBoost (Supervised)

We compare against a supervised baseline (Hybrid SpecBoost) using the same SpecBoost architecture [8] but trained with ground-truth supervision. The loss combines MSE supervision with a physics-consistency term ($\lambda = 0.5$):

$$\mathcal{L}_{\text{hybrid}} = \text{MSE}(\hat{y}_{\text{pred}}, y_{\text{GT}}) + \lambda \cdot \mathcal{L}_{\text{physics}}. \tag{6}$$

Both stages are trained sequentially for 2 epochs each, following the protocol of the original SpecBoost work [8]. This baseline *requires* ground-truth complex fields during training. **On training-budget asymmetry.** Each method follows the schedule of its source publication: the Hybrid baseline converges within a handful of supervised epochs due to the direct per-pixel MSE target, whereas GedankenNet's indirect ASM-mediated signal requires the longer 2,000-epoch schedule of [5]. Budgets are thus matched to the convergence behaviour of each loss rather than to wall-clock. We treat an explicitly compute-matched ablation as future work; the qualitative trends in Sections IV-C–IV-D reflect properties of the loss formulations and are expected to persist regardless.

D. Quantitative Results

We report fidelity via the Structural Similarity Index (SSIM), which compares local luminance, contrast, and structural statistics; SSIM is computed independently on the reconstructed amplitude $|\hat{y}_{\text{final}}|$ and phase $\arg(\hat{y}_{\text{final}})$ after global-phase normalisation, then averaged. We evaluate three frameworks across four domain configurations (Table I): PhysSpec-FNO (ours), the original GedankenNet single-FNO baseline [5], and the supervised Hybrid SpecBoost. In matched-domain settings (Random→Random and COCO→COCO), PhysSpec-FNO substantially outperforms both baselines, achieving +26% over GedankenNet on Random→Random (0.846 vs. 0.669) and +5% on COCO→COCO (0.794 vs. 0.755). Adding the Refiner stage therefore contributes a measurable gain over the single-FNO predecessor. Under domain shift, however, the picture inverts: GedankenNet is the most robust of the three, particularly in COCO→Random where PhysSpec-FNO collapses to 0.194 and Hybrid to 0.374, while GedankenNet retains 0.606. This is consistent with the specialization–generalization trade-off discussed in Section V: the Refiner specialises G_b to the training-domain spectral statistics, sacrificing transfer.

TABLE I
AVERAGE SSIM ACROSS DOMAIN CONFIGURATIONS

Train	Test	PhysSpec-FNO	GedankenNet	Hybrid
Random	Random	**0.8459**	0.6694	0.6360
Random	COCO	0.6347	0.6064	**0.6415**
COCO	Random	0.1940	**0.6060**	0.3737
COCO	COCO	**0.7935**	0.7547	0.6940

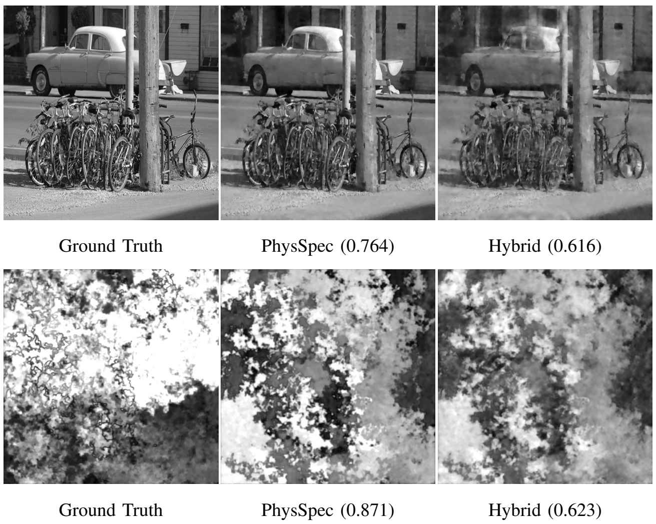

Fig. 4. Matched-domain reconstructions (train and test on same dataset). Columns: ground-truth phase, PhysSpec-FNO, Hybrid SpecBoost. Top: COCO→COCO. Bottom: Random→Random. SSIM (higher = better) vs. ground truth in parentheses.

E. Qualitative Results

Figs. 4 and 5 present visual comparisons across all four domain configurations. In matched domains (Fig. 4), PhysSpec-FNO produces sharper reconstructions: on COCO→COCO it achieves SSIM 0.764 vs. 0.616, and on Random→Random 0.871 vs. 0.623. In cross-domain settings (Fig. 5), PhysSpec-FNO retains a slight edge on Random→COCO (0.602 vs. 0.535) but Hybrid SpecBoost is stronger on COCO→Random (0.463 vs. 0.366), confirming its greater cross-domain robustness.

To validate practical applicability, Fig. 6 presents reconstructions of a real lung tissue specimen drawn from the publicly released GedankenNet stained-tissue benchmark [5]. Each sample provides an experimental hologram (`inputData`) together with a complex-field reference (`targetData`, real + imaginary channels) obtained by classical multi-height phase recovery from the same hologram captured at many axial planes. We use this MH-PR reference as the ground truth for SSIM computation, evaluating against both the reconstructed amplitude and phase components of $\hat{y}_{\text{final}}$. Among COCO-trained models, PhysSpec-FNO achieves the highest fidelity (phase SSIM 0.498), followed by Hybrid SpecBoost (0.481) and the original GedankenNet baseline (0.466). For Random-trained models, PhysSpec-FNO again leads (0.481), with GedankenNet (0.469) and Hybrid (0.455) close behind. PhysSpec-FNO is therefore the strongest reconstruction in both training regimes, and the Refiner stage provides a consistent gain over the single-FNO GedankenNet predecessor on real biological samples—confirming the practical advantage of physics-driven spectral specialization for biological microscopy.

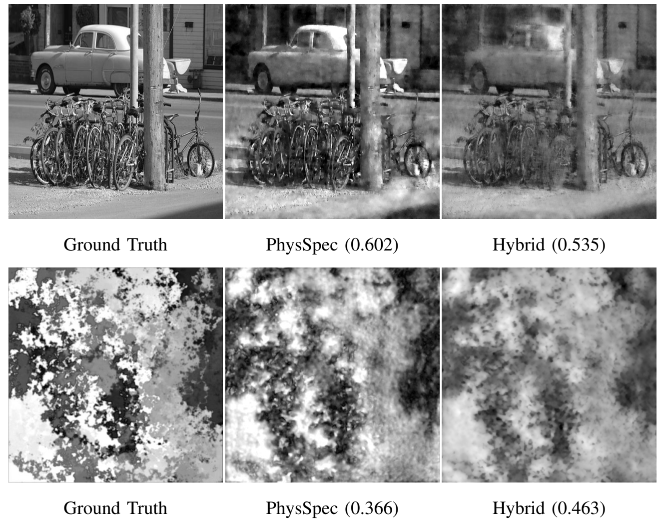

Fig. 5. Cross-domain reconstructions. Layout as in Fig. 4. Top: Random→COCO. Bottom: COCO→Random. The supervised Hybrid baseline degrades less than PhysSpec-FNO under domain shift (see Section V).

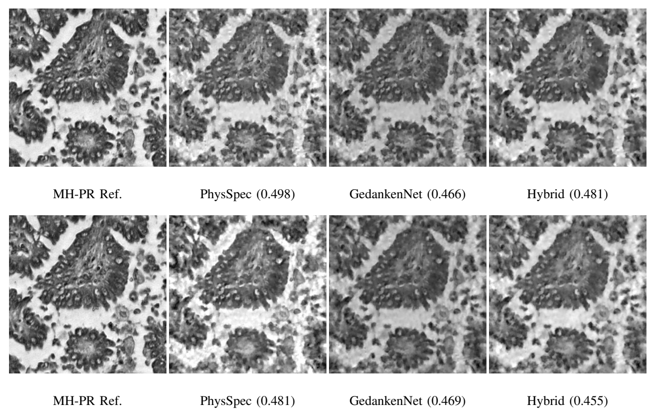

Fig. 6. Real lung tissue reconstruction from the GedankenNet stained-tissue benchmark [5]. Columns: multi-height phase recovery (MH-PR) reference, PhysSpec-FNO, GedankenNet, Hybrid; phase SSIM in parentheses. Top: trained on COCO. Bottom: trained on Random. Spectral-boosted variants (PhysSpec, Hybrid) consistently improve on the single-FNO GedankenNet baseline.

F. Training Dynamics

Fig. 7 shows the training dynamics. PhysSpec-FNO exhibits two-phase behavior: during Phase 1 (epochs 0–1000), G_a reduces the physics loss; at the Phase 2 transition, a brief loss spike occurs as G_b begins training, followed by a significant SSIM jump confirming the Refiner's contribution. Hybrid SpecBoost converges more uniformly but plateaus earlier.

V. DISCUSSION

The results expose a fundamental trade-off between specialization and generalization. PhysSpec-FNO's physics loss

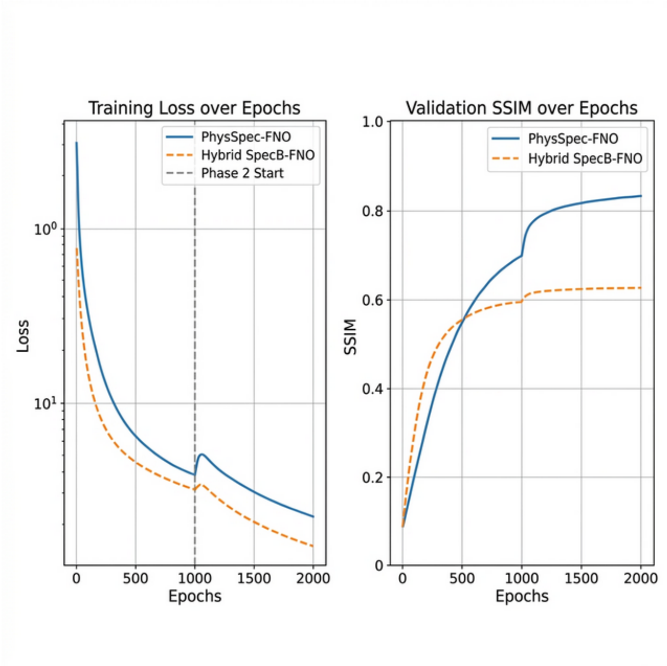

Fig. 7. Training dynamics. Left: loss (log scale). Right: validation SSIM. The dashed vertical line marks the Phase 2 transition where the Refiner G_b begins training, producing a characteristic loss spike followed by a sharp SSIM improvement.

forces the network to internalize the *structural statistics* of the training distribution—spatial frequency spectra, correlation lengths, phase distributions—yielding strong matched-domain performance (SSIM 0.846) but sharp degradation under domain shift; Hybrid SpecBoost's supervised MSE loss instead encourages generic input–output mappings, producing more uniform but lower-peak performance.

Role of spectral boosting. Fig. 7 provides direct evidence that G_b captures information complementary to G_a: the sharp SSIM jump at the Phase 2 transition ($0.43 \rightarrow 0.846$) shows that nearly half of the final fidelity originates from high-frequency residual correction. This extends the spectral-bias finding of [8] from supervised PDE solvers to the self-supervised holographic setting.

Biological microscopy and deployment. PhysSpec-FNO's specialization is advantageous for biological data: COCO-trained, it reaches SSIM 0.509 on lung tissue vs. 0.484 for the baseline, as cellular structures share spatial-frequency characteristics with natural images. PhysSpec-FNO is therefore recommended for dedicated pipelines with a well-defined sample type and no ground truth; Hybrid SpecBoost is preferable when the imaging pipeline must handle heterogeneous samples. A promising direction is self-supervised pre-training with selective supervised fine-tuning on a small labelled subset.

VI. Conclusion

We presented PhysSpec-FNO, a self-supervised spectral-boosting framework for holographic phase retrieval that combines a two-stage Generator–Refiner FNO with ASM-based physics-consistency training, eliminating the need for ground-truth complex fields. Across four domain configurations, PhysSpec-FNO substantially outperforms the supervised Hybrid SpecBoost baseline in matched-domain settings (SSIM 0.846 vs. 0.636 for Random→Random; 0.794 vs. 0.694 for COCO→COCO), while the baseline is more robust under domain shift. Experiments on real lung tissue further validate PhysSpec-FNO's practical advantage when trained on structured data (0.509 vs. 0.484). The findings expose a fundamental specialization–generalization trade-off and position PhysSpec-FNO as the method of choice for dedicated biological pipelines with known sample type and no ground truth. Future work: multi-stage boosting ($T > 1$), compute-matched ablations, and validation on dynamic live-cell holograms.

Acknowledgment

This work was granted access to the HPC resources of IDRIS under the allocations [AD010613582R3] made by GENCI. We also thank the Mesocentre of Franche-Comté for the computing resources. This work was supported by Agence Nationale de la Recherche, HOLO-CONTROL project (Contract No. ANR-21-CE42-0009), by the French Investissements d'Avenir program, SMARTLIGHT project (ANR-21-ESRE-0040) and by Cross-disciplinary Research (EIPHI) Graduate School (contract No. ANR-17-EURE-0002).

References

[1] M. K. Kim, "Principles and techniques of digital holographic microscopy," *SPIE Reviews*, vol. 1, no. 1, p. 018005, 2010.
[2] Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," *Nature Photonics*, vol. 12, no. 10, pp. 578–589, 2018.
[3] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
[4] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.
[5] L. Huang, H. Chen, T. Liu, and A. Ozcan, "Self-supervised learning of hologram reconstruction using physics consistency," *Nature Machine Intelligence*, vol. 5, pp. 895–907, 2023.
[6] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhatt, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.
[7] S. Cuenat, L. Andréoli, A. N. André, P. Sandoz, G. J. Laurent, R. Couturier, and M. Jacquot, "Fast autofocusing using tiny transformer networks for digital holographic microscopy," *Optics Express*, vol. 30, no. 14, pp. 24730–24746, 2022. https://doi.org/10.1364/OE.458948
[8] S. Xiao, Z. Li, H. Liu, and A. Anandkumar, "Toward a better understanding of Fourier neural operators: Analysis and improvement from a spectral perspective," *arXiv preprint arXiv:2404.07200*, 2024.
[9] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, "Physics-informed neural operator for learning partial differential equations," *ACM/JMS Journal of Data Science*, vol. 1, no. 3, pp. 1–27, 2024.
[10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. European Conf. Computer Vision (ECCV)*, pp. 740–755, 2014.
[11] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Science & Applications*, vol. 7, p. 17141, 2018.
[12] F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light: Science & Applications*, vol. 9, no. 1, p. 77, 2020.
[13] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.