Deep Coded Wavefront Sensing: Bridging the Simulation-Experiment Gap

Syed M. Kazim Patrick Müller Ivo Ihrke
Zentrum für Sensorsysteme
University of Siegen
Siegen, Germany 57076
{syed.kazim,patrick.mueller,ivo.ihrke}@uni-siegen.de

Abstract

Coded wavefront sensing (CWFS) is a recent computational quantitative phase imaging technique that enables one-shot phase retrieval of biological and other phase specimens. CWFS is readily integrable with standard laboratory microscopes and does not require specialized labor for its usage. The CWFS phase retrieval method is inspired by optical flow, but uses conventional optimization techniques. A main reason for this is the lack of publicly available datasets for CWFS, which prevents researchers from using deep neural networks in CWFS. In this paper, we present a forward model that utilizes wave optics to generate SynthBeads: a CWFS dataset obtained by modeling the complete experimental setup, including wave propagation through refractive index (RI) volumes of spherical microbeads, a standard microscope, and the phase mask, which is a key component of CWFS, with high fidelity. We show that our forward model enables deep CWFS, where pre-trained optical flow networks finetuned on SynthBeads successfully generalize to our SynthCell dataset, experimental microbead measurements, and, remarkably, complex biological specimens, providing quantitative phase estimates and thereby bridging the simulation-experiment gap.

1 Introduction

Biological cells consist of relatively non-absorbing structures with refractive indices (RI) close to that of water [1], making their imaging inaccessible through regular microscopy. However, internal biological structure is crucial to the research on cell growth, interaction, and development. For this purpose, we turn to quantitative phase imaging (QPI) techniques, which quantify light delay as it passes through transparent or weakly absorbing specimens [2, 3, 4]. Light passing through specimens is represented by a complex scalar wavefield, where the delay, referred to as the optical path difference (OPD), is encapsulated by the phase of the wavefield. The goal of QPI is to retrieve this phase, which is lost when the image sensor collects the wavefield. Over the years, several QPI techniques have been proposed, including interferometric methods that rely on the coherent interference of two wavefields [5, 6, 7], transport-of-intensity approaches that require measurements at multiple focal planes [8, 9], and computational phase retrieval methods that obtain multiple measurements while varying an optical element in the setup [10, 11].

Coded wavefront sensing (CWFS) is a modern single-shot QPI technique, that requires a single calibration image I_0 for an imaging system and a specimen image I_1 to retrieve the phase [12, 13]. CWFS hardware, which consists of a thin binary pattern attached to the sensor of a regular camera, is easily integrable with standard laboratory microscopes. CWFS is therefore an attractive QPI choice to image synthetic objects such as lenses and biological specimens such as cells with high accuracy and spatial resolution [14, 15]. The CWFS algorithm is a modified optical flow, which retrieves the phase by estimating the optical flow [16] between I_0 and I_1, even when there

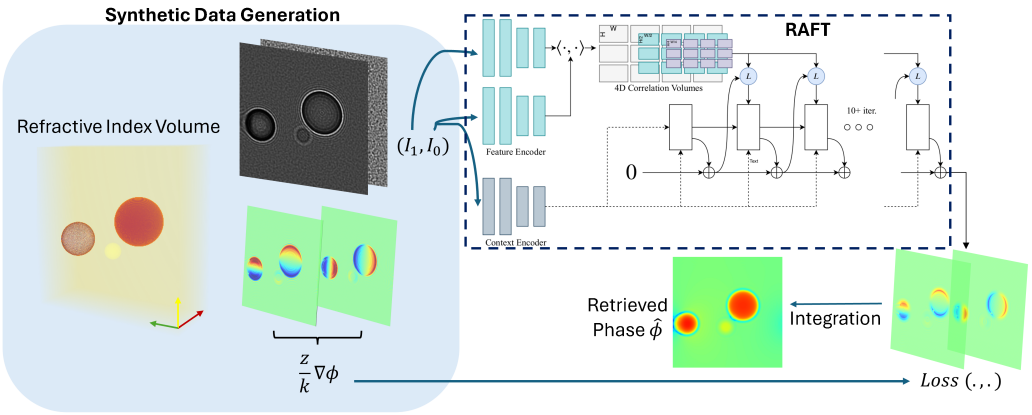

Figure 1: **DCWFS Workflow.** A plane wave traverses a 3D refractive index (RI) volume of randomly sampled spheres, producing a reference-specimen pair (I_0, I_1) of intensities which are provided to RAFT. RAFT estimates a gradient vector field, which is compared with the ground truth to calculate the loss and finetune the network. The gradient is integrated to retrieve the phase.

exists non-trivial brightness variation, by incorporating amplitude and curvature components in the minimization function [14, 17]. However, the formulation of this method is purely geometric and works in conditions when wave optical effects are not prevalent.

In recent years, deep neural networks (DNN) have replaced classical methods to compute optical flow in many domains such as robotics and automotive applications [18, 19, 20], resulting in impressive results [18, 19, 21]. However, this also requires high-quality image datasets and sufficient regularization with e.g. data augmentation such that the models can generalize to unseen data. In the specialized CWFS domain of microscopy, we are not aware of such a dataset, which prevents researchers and practitioners from leveraging the potential of modern DNN-based optical flow for CWFS. Moreover, due to the brightness changes between I_0 and I_1 and having speckle images, the CWFS data class is fundamentally different from other optical flow datasets.

In this paper, we propose a synthetic data generation pipeline to generate a realistic CWFS dataset, SynthBeads, to finetune a state-of-the-art DNN, RAFT [18], thereby enabling deep CWFS (DCWFS). Incorporating physics-based effects due to diffraction, limited numerical aperture, free-space wave propagation, and phase mask-induced speckles into the training of a DNN mitigates the simulation-experiment domain gap, which is observed as the degradation in performance when networks trained on artificial datasets fail to generalize to experimentally measured data.

In the proposed pipeline, we simulate wavefield propagation through a 3D RI volume of several randomly sampled spheres, followed by a microscope and finally a phase mask on the way to the image plane, where the camera records the intensity. We chose spheres because i) they can be simulated with high accuracy, ii) various configurations of spheres can be randomly sampled, reducing the risk of overfitting, and iii) we hypothesize that, locally, the spheres sufficiently approximate the optical flow vector fields corresponding to those of more complex specimens. Therefore, a network fine-tuned solely on spheres has the potential to generalize to other real-world specimens.

Fig. 1 shows the DCWFS workflow. The Synthetic Data Generation block summarizes the process by displaying the randomly sampled RI volume, the corresponding (I_0, I_1) pair, and the target flow fields. A pre-trained RAFT is finetuned using the SynthBeads dataset. The flow field estimated by the network is input to a Fourier integrator to retrieve the phase of the specimen.

Additionally, we construct an experimental microscopic setup in our laboratory, shown in Fig. 2 **a**, to collect CWFS data of a focal stack of real silica microbeads and acquire CWFS data of biological HEK cells from [15]. We demonstrate the generalization to unseen measured data of not only microbeads, but also complex biological HEK cells. Lastly, we compare the performance of DCWFS with an ADMM-based conventional CWFS phase retrieval method [14, 15, 22], both on synthetic and measured data.

2 Related work

Datasets and Learning The availability of high-quality training data is central in deep learning and enables models to robustly generalize to examples, which were not seen during training. Most importantly, the models need to infer on real-world data with high reliability and robustness in adverse conditions [23]. However, often there is a lack of suitable training data, especially for supervised learning. Different methods exist to overcome this limitation and their applicability for CWFS are discussed below.

A standard approach is to use data augmentation, which ranges from simple geometric augmentations on the training images to more advanced methods [24, 25, 26]. Developing such methods has received much attention [25], since these help in regularizing the model to avoid overfitting, increase the robustness and are usually combined with below methods. Other approaches include semi-supervised learning, where the model learns from a small initial set of annotated image data and makes use of non-labelled data, which is often more accessible [27]. The massive self-supervised pre-training of foundation models on unlabeled data [28, 29] points in a similar direction. This enables models to first learn general-purpose features, which can then be transferred to a specialized downstream domain for which only a comparably small amount of data is required [28]. This process is called transfer learning [30] and refers to transferring from a general-purpose model to the downstream task such as optical flow estimation or to the downstream domain such as CWFS. For the niche application of CWFS, we are neither aware of publicly available unlabeled nor labeled data that can be used effectively, which is why the above approaches still have a limited applicability for CWFS. With our dataset, we make a step towards closing this gap.

In addition to the above concepts, learning from and creating synthetic data is well-known in computer vision [31, 32, 33] and has powered the development of optical flow models with e.g. the FlyingChairs dataset [34] and its extension FlyingThings3D [35]. The idea is that the synthetic data mimics the real data's features by sampling from image generation models [36] or using simulation approaches from photo-realistic image rendering [31, 35, 37]. Generating synthetic data has many advantages such as limited costs, the ability to simulate rare or hard to measure cases. Importantly, balanced data can be sampled in a controlled setting as much as required, where variations are often easily achieved [33]. However, the quality and realism of the sampled synthetic data is crucial in ensuring that a properly regularized model learns the intended features and is able to generalize to real-world data. We refer to the discrepancy between the accuracy on synthetic and measured data as the simulation-experiment gap, which sometimes is also named the "reality gap" [37]. Our synthetic data generation process, which is based on a wave optics simulation for CWFS, provides the community with a means of enabling DNN training in the CWFS domain for the first time. We combine high quality synthetic data generation with transfer learning and demonstrate promising performance on measured data and contribute in closing the simulation-experiment gap.

Optical Flow Optical flow estimation is a long-standing task in computer vision for motion estimation. Classical approaches have formulated optical flow as a variational approach, where methods based on the Horn-Schunck approach dominated the field for a long time [38]. Inspired by the success of DNNs for image classification, enabling the potential of DNNs for optical flow triggered much research. It was FlowNet [34], which may have shifted the paradigm from energy-based methods, either pure or combined with feature learning, to models that do optical flow estimation in an end-to-end fashion [38]. Key to this promising approach was to overcome the lack of suitable training data by first creating a *synthetic* dataset for optical flow and then training the model on it [38]. Although trained on synthetic data, the method generalized to realistic test datasets [34]. Later, FlowNet2.0 [39] largely improved the initial model with changes in architecture and training method resulting in high accuracy and efficiency on standard test datasets such as MPI Sintel [40] or KITTI [20].

More recent approaches include RAFT [18], GMFlow [19] or diffusion model based optical flow estimation [21]. With foundation models available for various computer vision tasks, including depth estimation [41] and optical flow [21], another paradigm shift is ongoing towards large scale pre-training with massive computational power enabling impressive general zero-shot performance. However, they still often lack in accuracy in specialized domains and images from different modalities [28]. In this article, with limited computational resources available, we focus on a smaller state-of-the-art model, RAFT [18], with which we obtain quantitative results on microscopic speckle data

for CWFS for the first time. Still, we acknowledge that using more recent, larger foundation models for optical flow estimation may further enhance the accuracy.

3 Wavefront Sensing

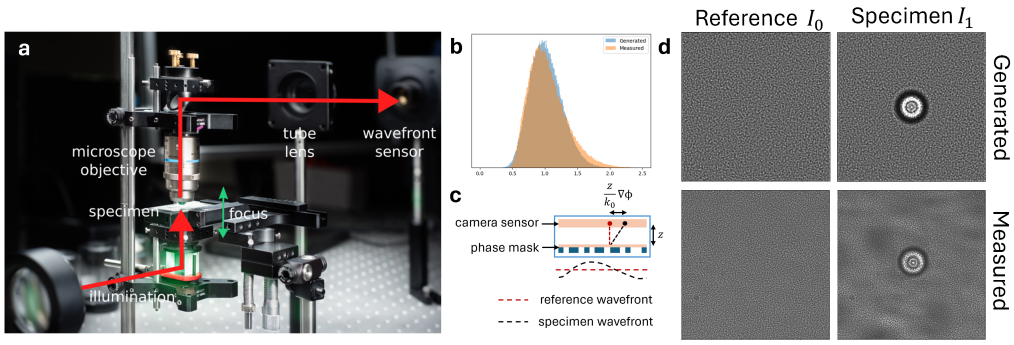

Figure 2: **a** Experimental setup of CWFS. **b** The histograms of synthetic and measured reference speckle patterns in (**d**). **c** Principle of CWFS. The movement of pixels is proportional to the gradient of the specimen wavefront in the phase mask plane and the distance z. **d** Synthetically generated and experimentally measured (I_0, I_1) pair corresponding to a sample of an out-of-focus microbead.

CWFS is a computational QPI technique. In CWFS, we measure a reference-specimen pair (I_0, I_1) of images [12, 13]. The reference is measured without the specimen, while the rest of the optical system, as shown in Fig. 2 **a**, remains unchanged. At the heart of the system is a standard microscope, where we attach a random phase mask to the cover glass of the image sensor. The phase mask is an optical element, usually with a binary phase pattern, that modifies the shape of the wavefront by delaying the incident wave by the same factor at random locations. The measurements I_0 and I_1 are both referred to as speckle patterns due to the bright and dark regions caused by the random delays.

The phase ϕ of the weakly-absorbing specimen is retrieved by analyzing the pair, (I_0, I_1). The relationship between the two is described by optical flow, where pixel displacements are proportional to the wavefront gradient in the phase mask plane, as seen in Fig. 2 **c**. Phase retrieval using CWFS is a two-step process: The first is essentially an *optical flow estimation*, formulated as,

$$I_1(r) = I_0(r + \frac{z}{k_0}\nabla\phi(r)), \tag{1}$$

where r denotes a coordinate point on the 2D plane, z is the distance of the phase mask from the sensor, and $k_0 = \frac{2\pi}{\lambda}$ is the wavenumber. The second step is simply the integration of the estimated gradient vector field $\nabla\phi(r)$ to obtain the phase.

The conventional optical flow problem assumes brightness constancy, i.e., the intensity value of the feature remains invariant as the pixel is displaced. However, specimen wavefronts with non-zero curvature bend light towards and away from certain regions. As the curvature becomes larger, brighter and darker regions become more pronounced in the speckle pattern, as seen around the silhouettes of the microbeads in Fig. 3. The changes in field magnitudes between the phase mask and the sensor are therefore due to the curvature of the wavefield. This effect is incorporated by including the Laplacian Δ of the phase in Eq. 1. Additionally, if the specimen absorbs light, or due to its volumetric RI distribution, light bends even for an "in-focus" image, appearing as intensity changes in I_1. The amplitude distribution A of the wavefield in the phase mask plane is also included to modify Eq. 1 to [14],

$$I_1(r) = A(r)(1 - \frac{z}{k_0}\Delta\phi(r))I_0(r + \frac{z}{k_0}\nabla\phi(r)). \tag{2}$$

As Eq. 2 is derived using the ray model of light, the accuracy of the relationship becomes tenuous as wave effects become prevalent. This can happen when z or $\Delta\phi$ increases to the point that the

condition $\frac{z}{k_0}|\Delta\phi| \ll 1$ is no longer satisfied [17]. Moreover, despite Eq. 2 retrieving the phase at every pixel, in practice, the recovered bandwidth is restricted to one-fifth of the bandwidth offered by the camera sensor, which is determined by its pixel pitch [13, 17].

4 Deep Coded Wavefront Sensing

4.1 Data generation

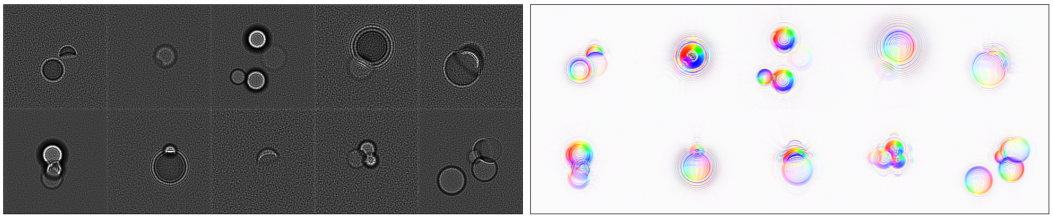

Figure 3: **Samples from SynthBeads.** I_1 and optical flow targets corresponding to RI volumes of randomly sampled spheres.

At the core of enabling DCWFS is the realistic synthetic data generation. We ensure that the features due to the wave optical nature of light, prevalent in microscopic data, are accurately represented in the data. Moreover, by simulating the entire CWFS pipeline, we implicitly incorporate the knowledge of the forward model into the training phase of the network, enabling it to bypass the shortcomings of model-blind methods. For each data point, we require a reference-specimen pair of intensities (I_0, I_1), along with the gradient vector field $\nabla = (\nabla_x, \nabla_y)$ of the specimen wavefield's phase in the phase mask plane, scaled by the distance from the sensor z.

The data generation pipeline is the following multi-step process: To generate the tuple $(I_0, I_1, \nabla)_m$ for $m \in \{1, \cdots, M\}$, where M is the total number of data points, we begin by discretizing the RI volume into a voxelized grid $\mathcal{R} = \{\boldsymbol{X}_j \mid j = 1, \cdots, J\}$, where $X_j \in \Omega \subseteq \mathbb{R}^3$. We define the refractive index over $\mathcal{R}$ at each iteration k for $k \in \{1, \cdots, K\}$ as $\eta^k(\boldsymbol{X}_j)$, where $K \sim \mathcal{U}(k_{\min}, k_{\max})$ is the number of spheres sampled for data point m. At each k, we sample the parameters of a sphere, namely the position vector of the center $\boldsymbol{C}_s$, RI η_s, and diameter d_s, from their joint probability distribution defined by the tensor product of independent random variables, respectively, as

$$(\boldsymbol{C}_s, \eta_s, d_s)^k \sim \mathcal{G}(\mu_{\boldsymbol{C}}, \sigma_{\boldsymbol{C}}) \otimes \mathcal{G}(\mu_\eta, \sigma_\eta) \otimes \mathcal{U}(d_{\min}, d_{\max}), \tag{3}$$

where $\mathcal{G}$ and $\mathcal{U}$ represent univariate Gaussian and uniform distributions, respectively. A mask corresponding to the sampled sphere, defined as

$$\boldsymbol{M}_j^k = \begin{cases} \mathbf{1}, & \|\boldsymbol{X}_j - \boldsymbol{C}_s\| \leq \dfrac{d_s}{2} \\ \mathbf{0}, & \text{otherwise} \end{cases}, \tag{4}$$

is used to update the RI voxels,

$$\eta^k(\boldsymbol{X}_j) = \eta^{k-1}(\boldsymbol{X}_j)(\mathbf{1} - \boldsymbol{M}_j^k) + \eta_s \boldsymbol{M}_j^k. \tag{5}$$

We initialize $\eta^0(\boldsymbol{X}_j) = \eta_0 = 1.518$ to simulate the RI of Zeiss Immersol (518 F), which is a standard material used to immerse specimens to reduce the RI contrast between the specimen and its surroundings. The final RI voxel grid $\eta^K(\boldsymbol{X}_j)$ for each i^{th} data point is split into $100\,\text{nm}$ thick slices along the optical axis, which is indicated by the red arrow in Fig. 1. We simulate a plane wave propagating along the optical axis that is modulated by each slice of the volume sequentially [42]. Modeling the propagation of light through a nonhomogeneous medium in this manner incorporates the effects of scattering and diffraction due to the RI contrast.

To introduce focus variations, the exit wavefield is propagated to a plane at a randomly sampled distance $d \sim \mathcal{G}(0, 4\,\mu\text{m})$ from the center of the voxel grid. Next, we model a standard microscope to map the wavefield to the image plane by cropping the spectrum of the wavefield corresponding to the numerical aperture of the objective lens, followed by a simple magnification step.

The remaining steps are exclusive to the image formation model of CWFS. In CWFS, a binary phase mask is attached to the image sensor (or to its cover glass). On its way to the image plane, the wavefield is modulated by the phase mask. The parameters of the CWFS phase mask, namely the delay t and the correlation ρ, chiefly determine the distribution of speckles, seen in Fig. 2 **b**. To ensure the DNN generalizes to a range of distributions, we model a new phase mask $\boldsymbol{P}$ after every 3^{rd} data point as follows,

$$(t, \rho) \sim \mathcal{U}(\frac{\lambda}{t_{\max}}, \frac{\lambda}{t_{\min}}) \otimes \mathcal{U}(\rho_{\min}, \rho_{\max}), \tag{6}$$

$$\boldsymbol{Q} \sim \mathcal{B}^{\otimes L}(p = 0.5), \tag{7}$$

$$\boldsymbol{P} = (\eta_0(1 - \boldsymbol{Q}) + \eta_{pm}\boldsymbol{Q}) * \mathcal{G}(0, \rho), \tag{8}$$

where $\mathcal{B}$ is the Bernoulli distribution, L is the total number of discrete structures of the phase mask, and $\eta_{pm} = 1.46$ is the RI of silica, which is used to fabricate the phase mask. An uncorrelated binary phase mask introduces large spatial frequencies, obscuring specimen features in the image plane. Therefore, in our simulations, the phase mask features are correlated by applying a Gaussian filter [43], mirroring the correlations introduced by the imperfect phase mask manufacturing process. The corresponding wavefield modulator is $\exp(jk_0\boldsymbol{P})$. More details are provided in the Appendix, Sect. A.1.

The scaled gradient of the wavefield, $\frac{z}{k_0}\nabla\phi$, is calculated in the phase mask plane, which becomes the target flow field for this data point, shown on the right in Fig. 3. The details of ground truth computation are provided in the Appendix, Sect. A.2. Lastly, the randomly delayed wavefield recorded in the image plane produces I_1, exemplified in Fig. 3. The same process is repeated without the 3D volume to produce the corresponding I_0. Gaussian noise with $\sigma = 1e^{-2}$ is added independently to I_0 and I_1, where the mean of I_0 is ≈ 1 as observed in Fig. 2 **b**.

4.2 Training

We finetune a pre-trained RAFT [18] model for DCWFS, without freezing any layers, on our synthetic microscopy dataset, which we refer to as SynthBeads. We use two strategies to generate data: First, we sample data on-the-fly (OTF) during the finetuning phase, which allows the use of $\mathrm{float32}$ intensity pairs (I_0, I_1) as input to the network, along with $\mathrm{float64}$ ground truth targets, without compression. We use a small mini-batch size of three and iterate for 1500 training steps. Second, we sample the data generator 4500 times to create a pre-sampled dataset (PSD). The data and corresponding targets are stored in the $\mathrm{uint16}$ format, saving time during finetuning since data generation alone takes $\approx 8\,\mathrm{h}$. Similarly to OTF, we iterate for 1500 training steps, randomly sampling three data points in each epoch. More details are provided in the Appendix, Sect. A.3.

5 Results

5.1 Synthetic data

To quantify the performance of the finetuned RAFT model and compare against the well-established ADMM-based CWFS algorithm (referred to simply as CWFS), we create two synthetic datasets. The datasets consist of 500 OTF-generated samples and are, therefore, input to the network without compression. The flow fields estimated by DCWFS are integrated to retrieve the OPDs.

We use the test split of SynthBeads as the first test dataset. The spheres are useful test specimens, as spherical silica microbeads are commonly imaged for research [44, 45, 46]. We provide a comparison of quantitative OPDs retrieved using DCWFS (trained with OTF and PSD samples) with CWFS, together with the corresponding ground truths in Fig. 4. The average l_2 loss for each method is shown in Tab. 1, which shows that both finetuned DCWFS outperform the standard CWFS and the pre-trained RAFT on the test split of SynthBeads.

We refer to the second as TestSynthCells, as it consists of samples with biological HEK cell-like features. TestSynthCells is produced by sampling a distribution of Zernike polynomials [47], which were fit to high-resolution phases of real HEK cells (see Appendix, Sect. B.1). Fig. 4 shows ground truth OPD samples and compares quantitative OPDs retrieved using DCWFS and CWFS. Although DCWFS generalizes well to synthetic cell samples, it struggles to distinguish fine structures, often

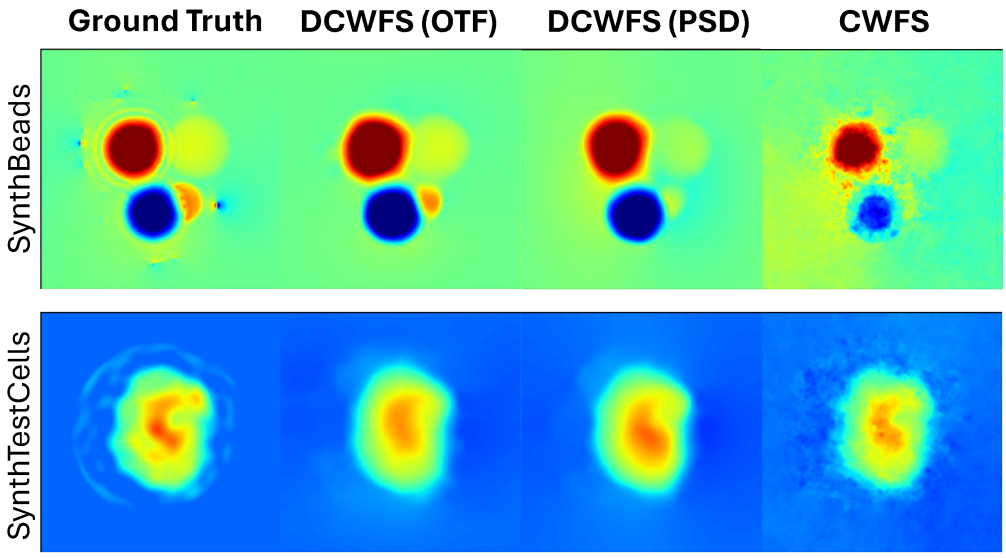

Figure 4: **Evaluations on synthetic test data.** Ground truth OPD and retrieved quantitative OPDs using DCWFS and the conventional CWFS optimization method.

Table 1: Performance Comparison on Synthetic Datasets (μm)

Dataset	DCWFS (Pretrained)	DCWFS (OTF)	DCWFS (PSD)	CWFS
SynthBeads (Test)	$0.68e^{-4}$	$\mathbf{0.20e^{-4}}$	$0.21e^{-4}$	$0.39e^{-4}$
TestSynthCells	$14.49e^{-4}$	$5.54e^{-4}$	$5.20e^{-4}$	$\mathbf{3.26e^{-4}}$

blending them together. The performance is summarised in Tab. 1, where finetuning yields substantial improvements over the pre-trained RAFT, but still falls short of CWFS. More evaluation results are provided in the Appendix, Sect.B.2.

5.2 Experimental data

Having trained the network to estimate optical flow on the SynthBeads dataset, we demonstrate the performance of the finetuned network on measured specimens. First, we establish the optical setup shown in Fig. 2 **a** to collect a focal stack of CWFS $10\,\mu\mathrm{m}$ silica beads, using a narrowband illumination source with FWHM of $\approx 5\,\mathrm{nm}$ centered at wavelength $\lambda = 526\,\mathrm{nm}$. The microbeads with RI $= 1.46$, immersed in Zeiss Immersol, are imaged using a 40x, 0.75 NA objective. The microscope slide, placed on a controlled z-stage, is displaced by $\pm 4\,\mu\mathrm{m}$ from its focus plane. Figs. 2 **b** and **d** provide a visual and quantitative comparison of the measured and synthetic data, where two important aspects are highlighted: First, the overlap of the histogram of the reference images verifies the similarity of the speckle distributions. Second, the simulated I_1 displays the features of the experimental microbead with high fidelity.

The measured focal stack of the silica microbeads is shown in Fig. 5, together with the OPDs estimated by DCWFS (OTF) and the conventional CWFS algorithm. Although both methods provide similar quantitative OPD maps, DCWFS OPD maps have fewer artifacts and cleaner boundaries compared to CWFS.

For a second set of measurement results, we evaluate the performance of DCWFS on complex biological specimens. We acquire the data captured by the authors in [15] of two instances of HEK cells. These are biological cancer cells with nonhomogeneous RI, ranging from 1.33 to 1.39, suspended in PBS and rotating in an acoustofluidic chamber [48]. The speckle pattern intensities of the HEK cells, together with the retrieved quantitative OPDs using DCWFS (OTF) and CWFS, are shown in Fig. 6. The results demonstrate that finetuning with SynthBeads not only reduces

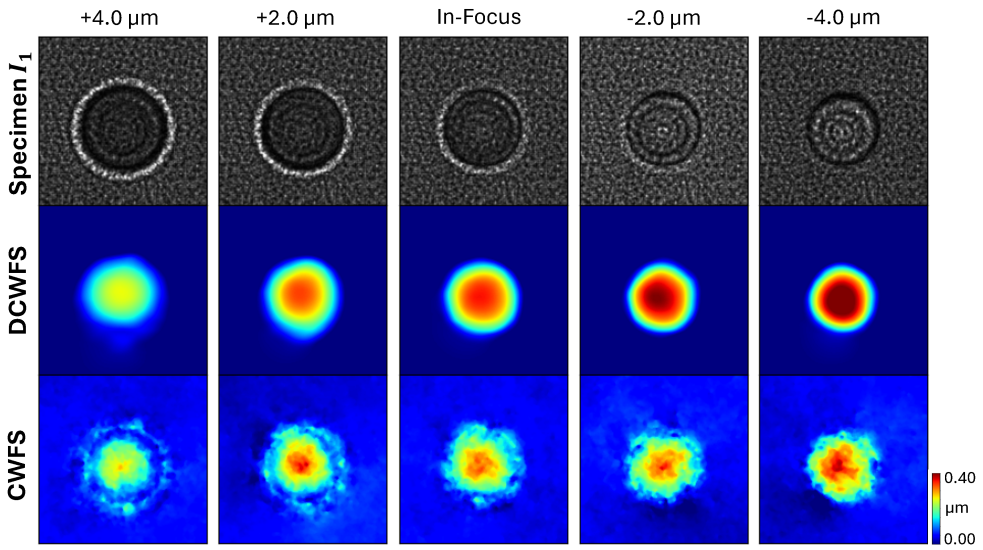

Figure 5: **Experimentally measured microbead focal stack ($\pm 4\,\mu\mathrm{m}$).** Intensity focal stack of a $10\,\mu\mathrm{m}$ silica bead, and the corresponding quantitative OPDs retrieved using DCWFS and the conventional CWFS optimization method.

the simulation-experiment gap, but also enables the network to generalize to complex real-world specimens. The results align with our hypothesis that the curvatures of the wavefields associated with spherical beads act as building blocks for more complex wavefields, indicating that training with various configurations of spherical beads can prove to be sufficient.

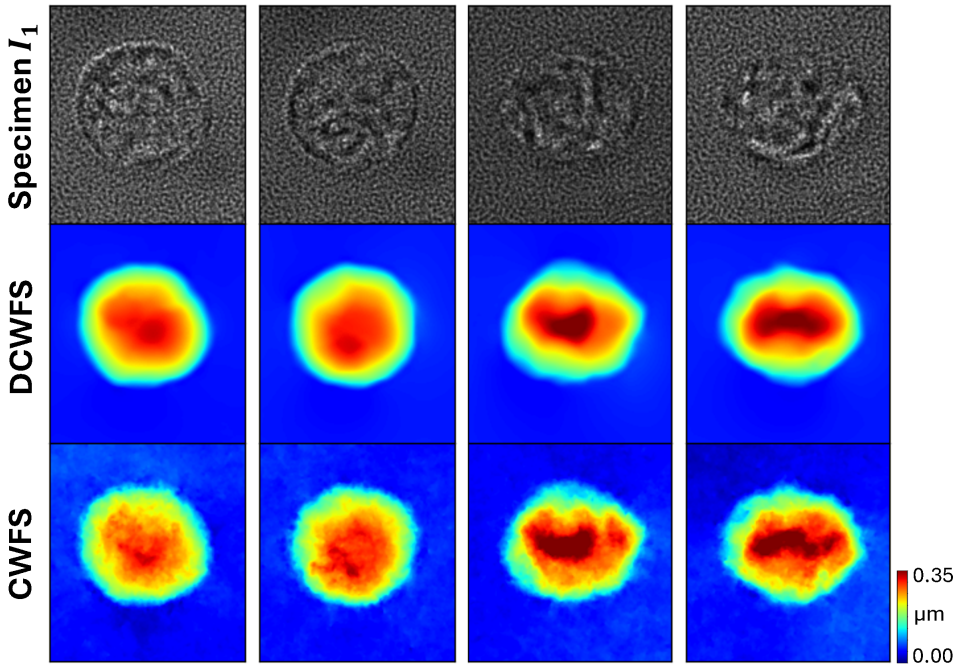

Figure 6: **Quantitative OPD of biological HEK cells.** I_1 of two different samples of complex cells and OPDs retrieved using DCWFS trained solely on SynthBeads, and the conventional CWFS optimization method.

6 Discussion and limitations

The forward model of CWFS proposed in this paper is modular. Given access to multiple 3D RI volumes of a specimen class of interest, the most effective approach would be to replace the microbead volumes with this specimen class.

However, one of the major hurdles in applying DNNs to niche physics applications is the lack of large publicly available datasets. In this paper, the performance of SynthBeads-finetuned RAFT demonstrates that sphere-induced local curvatures can serve as an effective basis set, analogous to radial basis functions, for reconstructing previously unseen complex wavefields. This suggests that the proposed framework is a powerful tool for modeling complex biological specimens without or with limited prior information.

The reconstruction improvement of DCWFS over CWFS, quantified on the test split of SynthBeads and qualitatively confirmed by the measured microbead results, demonstrates that the proposed CWFS modeling is well-suited to real-world data, narrowing the simulation-experiment gap. The success can be partly attributed to the wave optical simulation, as features resulting from diffraction and scattering, prevalent in microscopic data, are taken into account.

It can also be argued that the speckles in the CWFS data, contrary to being a nuisance, in fact facilitate generalization and are forgiving of inaccuracies in the forward model. I_0 is the diffraction pattern of the phase mask, which is compressed and stretched locally to produce I_1. Excluding the amplitude effects that may result from the absorption or thickness of the specimen, the optical flow estimation between I_0 and I_1 remains indifferent to the type of specimen, tracking primarily the features of the phase mask. Moreover, in simulating CWFS, the phase mask does not need to be the same as the one used experimentally, as long as the distributions of the experimental and simulated speckles overlap.

DCWFS seems to struggle with reconstructing fine structures, noticeable in the TestSynthCells OPD results. One way forward involves increasing the range of radii sampled, thereby including smaller beads in the finetuning phase. However, keeping the speckle distribution invariant, the speckles may obscure effects resulting from the small beads. Other directions could include augmenting the spheres with other elementary shapes, introducing effects due to edges, and elongated features that are otherwise not possible with spheres. Moreover, realistic aberrations can be added to distort the ideal wavefields using Zernike polynomials.

The current wave optical simulation works best when the phase mask is bandlimited, producing partially developed speckles, characterized by a low contrast speckle pattern with an approximately Gaussian distribution. The wavefield in the phase mask plane needs to be propagated over a few millimeters to the image plane, which becomes challenging without bandlimiting. Our measurements also feature partially developed speckles. However, the generalization of DCWFS may degrade for CWFS data with fully developed speckles.

Lastly, our simulation assumes coherent illumination, but CWFS itself does not have such an inherent limitation. Partial coherence modifies the size of the speckles and the distribution of the speckle pattern. Therefore, coherence must be incorporated into the simulation for comprehensive modeling.

7 Conclusion

CWFS is an easy-to-integrate computational imaging technique that retrieves high-resolution quantitative phase. The CWFS algorithm applies an optical flow-inspired conventional optimization method due to the unavailability of publicly available data. In this paper, we introduce SynthBeads, a synthetic dataset constructed by imaging 3D volumes of spherical microbeads. We provide a wave optics-based forward model that simulates the complete CWFS hardware pipeline to generate synthetic data. We show that finetuning a pre-trained DNN with SynthBeads enables DCWFS as a quantitative phase imaging technique.

We demonstrate quantifiable improvements over the conventional method on a test split of SynthBeads, and generalization to TestSynthCells, another synthetic dataset of HEK cell-like specimens. Lastly, we demonstrate that finetuning on SynthBeads i) narrows the simulation-experiment gap by evaluating the performance on experimentally measured beads, and ii) enables generalization to complex biological specimens by retrieving the quantitative phase of real biological cells.

8 Acknowledgements

We thank Monika Ritsch-Marte, Mia Kvåle Løvmo, and Simon Moser for establishing the cell-rotation measurement setup, and Judith Hagenbuchner and Michael Ausserlechner (Medical University Innsbruck) for providing us with the fixated HEK 293 cells. We further thank Wolfgang Heidrich for providing the coded wavefront sensing camera. This research was funded by Deutsche Forschungsgemeinschaft (IH 114/2-1, FOR 5336 Learning to Sense). Patrick Müller has additionally received funding through the ZESS-Scholarship from the NRW-Zentrum für Sensorsysteme.

References

[1] P. Y. Liu et al. "Cell Refractive Index for Cell Biology and Disease Diagnosis: Past, Present and Future". In: *Lab on a Chip* 16.4 (2016), pp. 634–644.

[2] R. V. Shack and B. C. Platt. "Production and use of a Lenticular Hartmann Screen." In: *Journal of the Optical Society of America A* 61 (1971), p. 656.

[3] Michael Reed Teague. "Deterministic Phase Retrieval: A Green's Function Solution". In: *Journal of the Optical Society of America* 73.11 (Nov. 1983), pp. 1434–1441. ISSN: 0030-3941. DOI: 10.1364/JOSA.73.001434.

[4] Etienne Cuche, Pierre Marquet, and Christian Depeursinge. "Spatial filtering for zero-order and twin-image elimination in digital off-axis holography". In: *Applied Optics* 39.23 (2000), pp. 4070–4075.

[5] Pierre Bon et al. "Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells". In: *Optics Express* 17.15 (2009), pp. 13080–13094.

[6] Emilio Sánchez-Ortiga et al. "Off-axis digital holographic microscopy: practical design parameters for operating at diffraction limit". In: *Appl. Opt.* 53.10 (2014), pp. 2058–2066.

[7] Bahram Javidi et al. "Roadmap on digital holography". In: *Optics Express* 29.22 (2021), pp. 35078–35118.

[8] Chao Zuo et al. "Noninterferometric single-shot quantitative phase microscopy". In: *Optics Letters* 38.18 (2013), pp. 3538–3541.

[9] Chao Zuo et al. "Transport of Intensity Equation: a Tutorial". In: *Optics and Lasers in Engineering* 135 (2020), p. 106187.

[10] Yicheng Wu, Manoj Kumar Sharma, and Ashok Veeraraghavan. "WISH: wavefront imaging sensor with high resolution". In: *Light: Science & Applications* 8.1 (2019), p. 44.

[11] Shaowei Jiang et al. "Wide-field, high-resolution lensless on-chip microscopy via near-field blind ptychographic modulation". In: *Lab on a Chip* 20.6 (2020), pp. 1058–1065.

[12] Congli Wang et al. "Ultra-High Resolution Coded Wavefront Sensor". In: *Optics Express* 25.12 (2017), pp. 13736–13746.

[13] Pascal Berto, Hervé Rigneault, and Marc Guillon. "Wavefront Sensing with a Thin Diffuser". In: *Optics Letters* 42.24 (Dec. 2017), pp. 5117–5120. ISSN: 0146-9592, 1539-4794. DOI: 10.1364/OL.42.005117.

[14] Congli Wang et al. "Quantitative Phase and Intensity Microscopy using Snapshot White Light Wavefront Sensing". In: *Scientific Reports* 9.1 (2019), p. 13795.

[15] Syed Muhammad Kazim et al. "Coded wavefront sensing for video-rate quantitative phase imaging and tomography: validation with digital holographic microscopy". In: *Optics Express* 33.12 (2025), pp. 25198–25209.

[16] Berthold KP Horn and Brian G Schunck. "Determining optical flow". In: *Artificial Intelligence* 17.1-3 (1981), pp. 185–203.

[17] Congli Wang et al. "Modeling Classical Wavefront Sensors". In: *Optics Express* 28.4 (2020), pp. 5273–5287.

[18] Zachary Teed and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow". In: *European Conference on Computer Vision*. Springer. 2020, pp. 402–419.

[19] Haofei Xu et al. "GMFlow: Learning Optical Flow via Global Matching". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, pp. 8121–8130. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Xu_GMFlow_Learning_Optical_Flow_via_Global_Matching_CVPR_2022_paper.html (visited on 08/15/2025).

[20] A Geiger et al. “Vision Meets Robotics: The KITTI Dataset”. In: *The International Journal of Robotics Research* 32.11 (Sept. 1, 2013), pp. 1231–1237. ISSN: 0278-3649. DOI: 10.1177/0278364913491297. URL: https://doi.org/10.1177/0278364913491297 (visited on 09/08/2023).

[21] Saurabh Saxena et al. “The Surprising Effectiveness of Diffusion Models for Optical Flow and Monocular Depth Estimation”. In: *Advances in Neural Information Processing Systems* 36 (Dec. 15, 2023), pp. 39443–39469. URL: https://proceedings.neurips.cc/paper_files/paper/2023/hash/7c119415672ae2186e17d492e1d5da2f-Abstract-Conference.html (visited on 08/13/2025).

[22] Stephen Boyd et al. “Distributed optimization and statistical learning via the alternating direction method of multipliers”. In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.

[23] Dan Hendrycks and Thomas Dietterich. “Benchmarking Neural Network Robustness to Common Corruptions and Perturbations”. In: *Proceedings of the International Conference on Learning Representations 2019*. International Conference on Learning Representations. 2019. URL: https://openreview.net/forum?id=HJz6tiCqYm (visited on 11/17/2023).

[24] Sangdoo Yun et al. “CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features”. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, pp. 6023–6032. URL: https://openaccess.thecvf.com/content_ICCV_2019/html/Yun_CutMix_Regularization_Strategy_to_Train_Strong_Classifiers_With_Localizable_Features_ICCV_2019_paper.html (visited on 11/13/2023).

[25] Dan Hendrycks et al. “The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization”. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, pp. 8340–8349. URL: https://openaccess.thecvf.com/content/ICCV2021/html/Hendrycks_The_Many_Faces_of_Robustness_A_Critical_Analysis_of_Out-of-Distribution_ICCV_2021_paper.html (visited on 11/13/2023).

[26] Dan Hendrycks* et al. “AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty”. In: International Conference on Learning Representations. 2020. URL: https://openreview.net/forum?id=S1gmrxHFvB (visited on 08/19/2024).

[27] Kihyuk Sohn et al. “FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence”. In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 596–608. URL: https://proceedings.neurips.cc/paper/2020/hash/06964dce9addb1c5cb5d6e3d9838f733-Abstract.html (visited on 08/15/2025).

[28] Rishi Bommasani et al. *On the Opportunities and Risks of Foundation Models*. July 12, 2022. DOI: 10.48550/arXiv.2108.07258. arXiv: 2108.07258 [cs]. URL: http://arxiv.org/abs/2108.07258 (visited on 08/14/2025). Pre-published.

[29] Maxime Oquab et al. “DINOv2: Learning Robust Visual Features without Supervision”. In: *Transactions on Machine Learning Research* (July 26, 2023). ISSN: 2835-8856. URL: https://openreview.net/forum?id=a68SUt6zFt (visited on 01/23/2024).

[30] Sinno Jialin Pan and Qiang Yang. “A Survey on Transfer Learning”. In: *IEEE Transactions on Knowledge and Data Engineering* 22.10 (Oct. 2010), pp. 1345–1359. ISSN: 1558-2191. DOI: 10.1109/TKDE.2009.191. URL: https://ieeexplore.ieee.org/document/5288526/?arnumber=5288526 (visited on 07/16/2024).

[31] German Ros et al. “The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 3234–3243. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.352. URL: http://ieeexplore.ieee.org/document/7780721/ (visited on 06/29/2022).

[32] Yohann Cabon, Naila Murray, and Martin Humenberger. *Virtual KITTI 2*. Jan. 29, 2020. DOI: 10.48550/arXiv.2001.10773. arXiv: 2001.10773 [cs]. URL: http://arxiv.org/abs/2001.10773 (visited on 08/15/2025). Pre-published.

[33] Goran Paulin and Marina Ivasic-Kos. "Review and Analysis of Synthetic Dataset Generation Methods and Techniques for Application in Computer Vision". In: *Artificial Intelligence Review* 56.9 (Sept. 1, 2023), pp. 9221–9265. ISSN: 1573-7462. DOI: 10.1007/s10462-022-10358-3. URL: https://doi.org/10.1007/s10462-022-10358-3 (visited on 08/14/2025).

[34] Alexey Dosovitskiy et al. "FlowNet: Learning Optical Flow With Convolutional Networks". In: Proceedings of the IEEE International Conference on Computer Vision. 2015, pp. 2758–2766. URL: https://openaccess.thecvf.com/content_iccv_2015/html/Dosovitskiy_FlowNet_Learning_Optical_ICCV_2015_paper.html (visited on 08/14/2025).

[35] Nikolaus Mayer et al. "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 4040–4048. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.438. URL: http://ieeexplore.ieee.org/document/7780807/ (visited on 08/15/2025).

[36] Zijie J. Wang et al. *DiffusionDB: A Large-scale Prompt Gallery Dataset for Text-to-Image Generative Models*. July 6, 2023. DOI: 10.48550/arXiv.2210.14896. arXiv: 2210.14896 [cs]. URL: http://arxiv.org/abs/2210.14896 (visited on 08/15/2025). Pre-published.

[37] Josh Tobin et al. "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World". In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Sept. 2017, pp. 23–30. DOI: 10.1109/IROS.2017.8202133. URL: https://ieeexplore.ieee.org/document/8202133/ (visited on 08/15/2025).

[38] Junhwa Hur and Stefan Roth. "Optical Flow Estimation in the Deep Learning Age". In: *Modelling Human Motion: From Human Perception to Robot Design*. Ed. by Nicoletta Noceti, Alessandra Sciutti, and Francesco Rea. Cham: Springer International Publishing, 2020, pp. 119–140. ISBN: 978-3-030-46732-6. DOI: 10.1007/978-3-030-46732-6_7. URL: https://doi.org/10.1007/978-3-030-46732-6_7 (visited on 08/14/2025).

[39] Eddy Ilg et al. "FlowNet 2.0: Evolution of Optical Flow Estimation With Deep Networks". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 2462–2470. URL: https://openaccess.thecvf.com/content_cvpr_2017/html/Ilg_FlowNet_2.0_Evolution_CVPR_2017_paper.html (visited on 06/26/2024).

[40] Daniel J. Butler et al. "A Naturalistic Open Source Movie for Optical Flow Evaluation". In: *Computer Vision – ECCV 2012*. Ed. by Andrew Fitzgibbon et al. Berlin, Heidelberg: Springer, 2012, pp. 611–625. ISBN: 978-3-642-33783-3. DOI: 10.1007/978-3-642-33783-3_44.

[41] Lihe Yang et al. "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024, pp. 10371–10381. URL: https://openaccess.thecvf.com/content/CVPR2024/html/Yang_Depth_Anything_Unleashing_the_Power_of_Large-Scale_Unlabeled_Data_CVPR_2024_paper.html (visited on 06/23/2024).

[42] J Van Roey, J Van der Donk, and PE Lagasse. "Beam-propagation method: analysis and assessment". In: *Journal of the Optical Society of America* 71.7 (1981), pp. 803–810.

[43] Joseph W Goodman. *Simulating speckle with Python*. SPIE Press, 2024.

[44] Yongjin Sung et al. "Optical diffraction tomography for high resolution live cell imaging". In: *Optics Express* 17.1 (2009), pp. 266–277.

[45] Ulugbek S. Kamilov et al. "Learning approach to optical tomography". In: *Optica* 2.6 (May 2015), p. 517. ISSN: 2334-2536. DOI: 10.1364/optica.2.000517. URL: http://dx.doi.org/10.1364/OPTICA.2.000517.

[46] Kevin C Zhou and Roarke Horstmeyer. "Diffraction tomography with a deep image prior". In: *Optics Express* 28.9 (2020), pp. 12872–12896.

[47] Hansol Kim et al. "Noise robust Zernike phase retrieval via learning based algorithm only with 2-step phase shift measurements". In: *Optics Express* 31.19 (2023), pp. 30248–30266.

[48] Mia Kvåle Løvmo et al. "Controlled Orientation and Sustained Rotation of Biological Samples in a Sono-Optical Microfluidic Device". In: *Lab on a Chip* 21.8 (2021), pp. 1563–1578.

A Finetuning Details

A.1 Data generation

- Data generation parameters:
  1. Object space resolution = $100\,\mathrm{nm} \times 100\,\mathrm{nm} \times 100\,\mathrm{nm}$
  2. $\mu_{\boldsymbol{C}} = (50\,\mu\mathrm{m}, 50\,\mu\mathrm{m}, 20\,\mu\mathrm{m})$, $\sigma_{\boldsymbol{C}} = (5\,\mu\mathrm{m}, 5\,\mu\mathrm{m}, 2\,\mu\mathrm{m})$
  3. $J = 500 \times 500 \times 200$
  4. $k_{min} = 1, k_{\max} = 5$
  5. $\mu_\eta = 1.5, \sigma_\eta = 0.07$
  6. $d_{min} = 6\,\mu\mathrm{m}, d_{\max} = 16\,\mu\mathrm{m}$
  7. Microscope: NA = 0.85, magnification = 60
  8. $t_{min} = 2, t_{\max} = 4$
  9. $\rho_{min} = 3, \rho_{\max} = 4$
  10. Image space resolution = $3\,\mu\mathrm{m} \times 3\,\mu\mathrm{m}$
  11. $L = 166 \times 166$ with physical structure size of $10\,\mu\mathrm{m} \times 10\,\mu\mathrm{m}$ each.
  12. $z \sim \mathcal{U}(1\,\mathrm{mm}, 3\,\mathrm{mm})$
- To store pre-sampled data, we convert the reference-spacimen pair to $\mathrm{uint16}$ by multiplying the pair by $1e^4$ and taking the int. The gradient fields are converted to $\mathrm{uint16}$ by applying $\min\max$ normalization, setting $\min = -32$ and $\max = 32$, multiplying the results by $65 \times 1e^3$, and taking the int.

A.2 Computing the ground truth

In the following, we outline the step-by-step procedure for computing the ground truth for data points in the SynthBeads and TestSynthCells datasets.

1. Calculate the difference between adjacent pixels along the x- and y-axes of the wavefield's phase or OPD in the phase mask plane.
2. Convert the resulting vector field to the gradient with physical units by dividing it by the resolution of the lateral plane in object space, p_x.
3. The complex wave field with unit amplitude is given by $e^{j\phi}$, where phase ϕ can be written in terms of OPD as $\phi = \frac{2\pi}{\lambda}\mathrm{OPD}$.
4. Remove the 2π peaks in the gradient vector field caused by phase wrapping by applying a median filter with a kernel size of 3.
5. Scale the vector field by $\frac{z}{k_0}$.
6. Finally, convert to pixel shifts by dividing once more by p_x to obtain the ground truth.
7. For QPI, the estimate of the network therefore needs to be scaled by $\frac{p_x^2}{z}$ before (or after) integrating to retrieve the phase.

A.3 Network details

We make a single change to RAFT [18]: We modify the first-layer convolution kernels of both the feature and context encoders from three channels to a single channel, averaging the corresponding weights to accommodate single-channel CWFS data, as shown in Fig. 7.

The hyperparameters and loss functions used in this paper are duplicated from [18], with more details provided below.

- RAFT large imported from Pytorch, pre-trained on FlyingChairs and FlyingThings3D with finetuning on the Sintel dataset.
- Loss function:

$$\mathcal{L} = \sum_{i=1}^{N} \gamma^{N-i} \left\| f_{gt} - f_i \right\|_1 ,$$

```python
# model
def init_model_RAFT(model, device='cuda', checkpoint=None):

    if checkpoint == None:
        print('Loading models with pretrained weights')
        weights = Raft_Large_Weights.DEFAULT.get_state_dict()
        weights['feature_encoder.convnormrelu.0.weight'] = torch.mean(weights['feature_encoder.convnormrelu.0.weight'], dim=1).unsqueeze(1)
        weights['context_encoder.convnormrelu.0.weight'] = torch.mean(weights['context_encoder.convnormrelu.0.weight'], dim=1).unsqueeze(1)
    else:
        print(f'Loading models with checkpoint: {checkpoint}')
        weights = torch.load(checkpoint, map_location=torch.device(device))

    # model = raft_large(progress=False)
    model.feature_encoder.convnormrelu[0] = torch.nn.Conv2d(1, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3))
    model.context_encoder.convnormrelu[0] = torch.nn.Conv2d(1, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3))

    model.load_state_dict(weights)

    return model.to(device)
```

Figure 7: **Modification to RAFT.**

where $\gamma = 0.8$ and $N = 12$ is the number of recurrent updates of the *update operator* of RAFT.

- The optimizer used is Adam, learning rate= $1e^{-4}$, weight decay=$1e^{-5}$.
- The generated data have an image size of 500×500, which is resized to match RAFT's input image size of 520×960.
- Both I_0 and I_1 are normalized by their respective means before being input to the network.
- Both OTF and PSD DCWFS finetuning is done on a single V100.
- OTF finetuning takes $\approx 8\,\mathrm{h}$ and $7\,\mathrm{min}$, whereas PSD involves $\approx 8\,\mathrm{h}$ pre-computation, followed by $\approx 27\,\mathrm{min}$ for finetuning. Since the pre-computation has to be done only once, its costs can be amortized quickly.

Lastly, to ensure that the neural network's output reflects the quantitative phase, we calibrate the model by scaling the output with a constant factor α. To estimate α, we evaluated the fine-tuned NN and the conventional CWFS algorithm on samples from the SynthBeads dataset, performed a pixel-wise division, and took the average. In our study, $\alpha \approx \frac{1}{\pi}$. We note that calibration requires a true ground truth, which can also be calculated analytically for simple specimens with weak RI contrast, and more generally, with advanced phase imaging techniques such as DHM.

B Synthetic Data

B.1 Synthetic HEK cell generation

Here, we require a few samples of the phases (possibly retrieved using another modality such as digital holographic microscopy (DHM) or available online) of the class of objects we are interested in. For example, if we want the network to predict the phases of HEK cells from Coded WFS-based speckle patterns, we would need phase images of HEK cells. For this paper, we acquire phases of the frames of a rotating HEK cell from the authors in [15]. We randomly select six frames and fit the first 860 Zernike polynomials separately to each phase. Given the set of six coefficients for each polynomial, the mean and standard deviation are calculated, providing 860 distributions. We sample these distributions to generate synthetic HEK cell OPDs. The standard deviations have been scaled by a factor of 5, and Gaussian noise with $\sigma = 5e^{-4}$ is added to the mean to diversify synthetic HEK specimens. Random crops, flips, and rotations are also added to produce the OPDs in Fig. 8. These OPDs are converted to wavefields in the phase mask plane, followed by the steps outlined in Sect. 4.1 to produce the corresponding specimen intensities I_1, also shown in Fig. 8.

B.2 Results on synthetic data

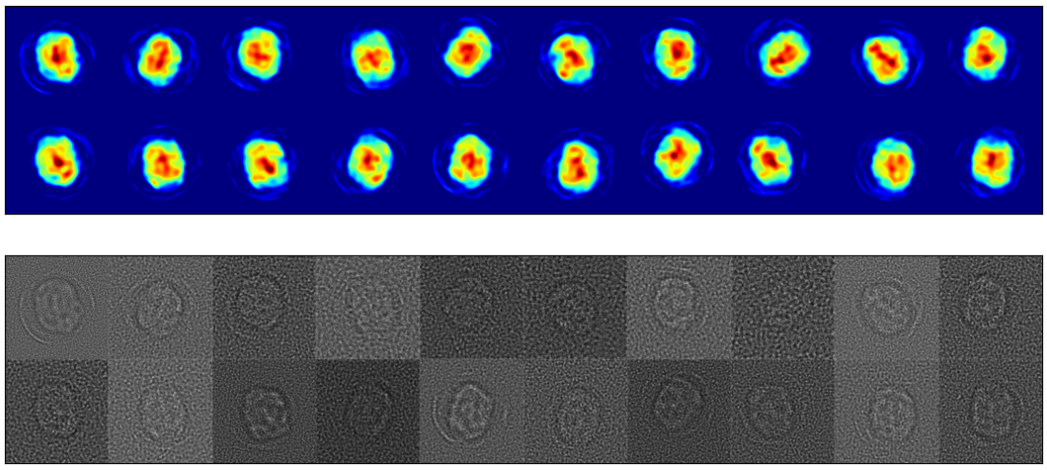

Figure 8: **Examples of TestSynthCells.** (Top) OPDs of the synthetic cells are generated by sampling the distributions of the Zernike coefficients. (Bottom) Specimen intensities I_1 are simulated by modulating the corresponding OPDs above by a random phase mask and propagating by a distance z to the image plane.

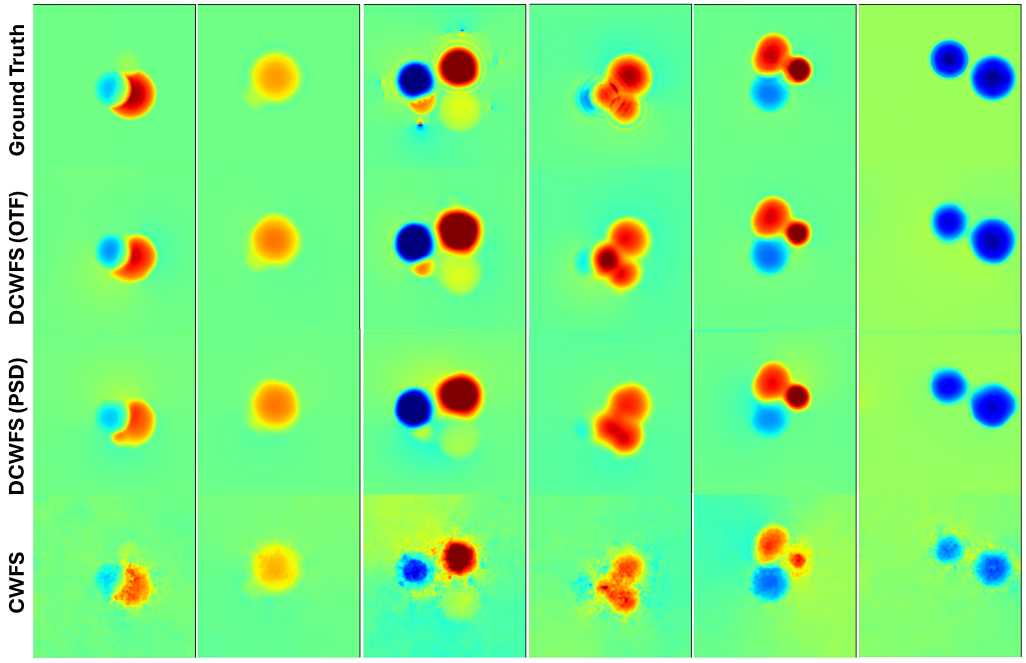

Figure 9: **Test split of SynthBeads**. Ground truth OPD and retrieved quantitative OPDs using DCWFS and the conventional CWFS optimization method.

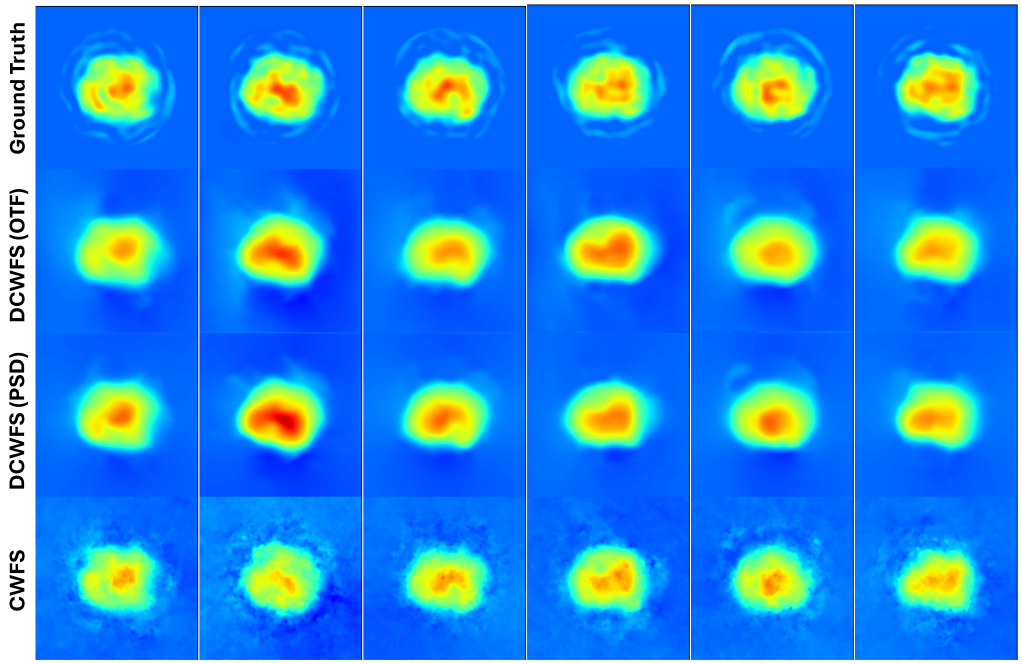

Figure 10: **TestSynthCells.** Ground truth OPD and retrieved quantitative OPDs using DCWFS and the conventional CWFS optimization method.