

HoloADMM: High-Quality Holographic Complex Field Recovery

Mazen Mel¹ , Paul Springer², Pietro Zanuttigh¹ , Zhou Haitao³, and Alexander Gatto²

¹ University of Padova, Department of Information Engineering, Padova, Italy
`{melmazen,zanuttigh}@dei.unipd.it`

² Sony Semiconductor Solutions Europe, Stuttgart Laboratory 1, Germany
`{paul.springer,alexander.gatto}@sony.com`

³ École polytechnique fédérale de Lausanne, Lausanne, Switzerland
`haitao.zhou@epfl.ch`

Abstract. Holography enables intriguing microscopic imaging modalities, particularly through Quantitative Phase Imaging (QPI), which utilizes the phase of coherent light as a way to reveal the contrast in transparent and thin microscopic specimens. Despite the limitation of image sensors, which detect only light intensity, phase information can still be recorded within a two-dimensional interference pattern between two distinct light waves. Numerical reconstruction is later needed to retrieve the amplitude and phase from such holographic measurements. To this end, we introduce HoloADMM, a novel interpretable, learning-based approach for in-line holographic image reconstruction. HoloADMM enhances imaging capability with spatial image super-resolution, offering a versatile framework that accommodates multiple illumination wavelengths and supports extensive refocusing ranges with up to 10 μm precision. Our results indicate a substantial improvement in reconstruction quality over existing methods and demonstrate HoloADMM’s effective adaptation to real holographic data captured by our Digital in-line Holographic Microscope (DIHM). This work not only advances holographic imaging techniques but also broadens the potential for non-invasive microscopic analysis applications.

Keywords: In-line holography, lens-free microscopy, ADMM, super-resolution, quantitative phase imaging.

1 Introduction

Dennis Gabor introduced holography in his seminal work [13] where he demonstrated true three dimensional imaging capability by simultaneously recording coherent light’s intensity and direction, i.e., phase, within a two dimensional interference pattern. Gabor’s prototype, depicted in Fig. 1 on the right, embodies an in-line holographic imaging setup. Here, an object of interest is illuminated by a coherent light source, generating a diffraction pattern at the detector plane obtained through the interference of two waves: one scattered by the object and

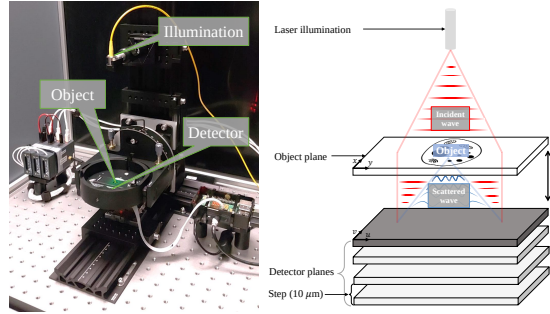


Fig. 1: Lens-free in-line holographic setup: (*Left*) our Digital In-line Holographic Microscope (DIHM). (*Right*) schematics of the different DIHM components.

another unobstructed background wave passing through the object plane. This interference pattern encodes both amplitude and phase information of the wave scattered by the object, thus characterizing its complete complex transmission distribution, which can be recovered using phase retrieval-based reconstruction techniques. The ability to capture phase information allows for interesting applications such as phase imaging in microscopy [24], where it allows to generate image contrast for transparent ultra-thin microscopic specimens. Advancements in phase imaging have facilitated close-up non-invasive inspection of living cells [28], with applications in medicine [29], biology [20], and neuroscience [8]. Highly accurate interferometry, driven by holographic imaging, finds applications in high-precision engineering [36] and material science. Phase contrast microscopy [2] also offers phase imaging capability but lacks quantitative measurement ability. In contrast, holography is able to quantify exactly the light phase shift making it suitable for applications requiring precise measurements, such as accurate tolerance estimations of microscopic features and three-dimensional object reconstruction [41], additionally, in-line holographic setups can be used in compact, mobile, and lensless imaging systems free of any optical aberrations. Despite its desirable features, in-line holography typically requires iterative numerical reconstruction [14] to retrieve high-quality images and to suppress undesirable artifacts, such as the *twin image* [39], which corresponds to the latent field's complex conjugate recorded as a byproduct by the image sensor and appears as an out-of-focus, blurry image superimposed onto the true sharp one.

Recent advancements have demonstrated the efficacy of learning-based methods in tackling holographic reconstruction, both in supervised [3, 4, 31] and unsupervised [19] manners and to suppress, to some degree, the twin image artifacts. The scarcity of real large-scale holographic measurements, coupled with ground truth complex transmission data, has led most approaches in the literature to rely primarily on synthetic data. However, due to the significant domain gap between synthetic and real holographic measurements, achieving model transferability to different domains remains challenging and no approach with effective generalization capabilities have been demonstrated so far.

We address for the first time this issue and propose a comprehensive and versatile framework based on an unrolled deep architecture inspired by a model-based reconstruction strategy that is robust to domain changes. This is made possible through our proposed model’s interpretability: in this way we can leverage large scale synthetic data to effectively learn the inverse holographic image formation model and showcase outstanding generalization ability to the real world domain without any explicit adaptation. Furthermore, we introduce a joint framework for in-line holographic image reconstruction and spatial image super-resolution where we explicitly exploit complimentary spatial information in the form of aliasing introduced by sub-pixel displacements between consecutive holographic measurements. Our main contributions are the following:

- This work introduces a versatile framework that seamlessly integrates in-line holographic image reconstruction with spatial super-resolution and supports an extensive refocusing range.
- Large scale synthetic data is leveraged to learn the inverse model underlying in-line holographic image formation within an interpretable manner which ensures robust generalization capabilities well beyond the data distribution of the training domain.
- Our approach is validated on standard datasets and also through real world samples imaged with our Digital in-line Holographic Microscope (DIHM): it achieves high-quality reconstruction and demonstrate the framework’s efficacy and practical utility in real-world applications.

2 Related Works

Iterative phase retrieval algorithms based on error reduction are commonly used in digital in-line holography to reconstruct the latent complex field from real-valued holographic measurements. Gerchberg and Saxton [14] first proposed an alternating field projections approach between the object and detector planes while enforcing support constraints. With sufficient number of iterations the field typically converges towards the latent one. Fienup [10] later introduced modifications to the original method of [14] notably the Hybrid Input Output (HIO) variant which incorporates a feedback parameter to relax the hard support constraint on the object plane. Indeed, it can be shown that such error reduction approaches are special cases of an inverse problem solving framework [25] prone to local minima and in some cases divergence. Fienup’s modification helps avoid possible stagnation and further regularizes the solution space, it has been widely used in the literature and it serves as a baseline method, with variants using single or multiple input holograms captured at different heights referred to as "SH-PR" or "MH-PR", respectively. Iterative methods leveraging prior knowledge of the target sample, such as sparsity [1, 9, 39], have demonstrated the ability to produce cleaner images by enforcing such constraints. Moreover, iterative solvers help mitigate the undesirable artifacts of the *twin image*, as demonstrated in previous studies. For instance, Zhang et al. [39] devised a compressive sensing

approach employing an iterative shrinkage/thresholding strategy to gradually reduce such artifacts. Latychevskaia et al. [21] achieved twin image-free in-line holography through an iterative alternating projections approach without enforcing prior knowledge on the object of interest. Chen et al. [6] proposed an iterative method for holographic image reconstruction and computational refocusing optimizing a least squares problem with plain gradient descent steps. Niknam et al. [27] and Chen et al. [38] employed an untrained neural network with randomly generated weights as a natural image prior within a model-based reconstruction framework. Adversarial iterative techniques have also been investigated in [7] where the authors used a generative network to learn the inverse image formation model of in-line holography, a discriminator is then used to distinguish between the original hologram and the simulated one.

Learning-based methods offer inherent immunity to the *twin image* issue. Networks trained using pairs of holographic images and sharp ground truth data learn a direct mapping between the two sets, circumventing the need to model the underlying physics of holographic image formation. Once trained successfully, these methods can produce twin image-free and sharp reconstructions. However, this straightforward black box approach may encounter unexpected failure cases when presented with new data exhibiting different statistics than that seen during training. Recent work by Chen et al. [3,4] demonstrated promising model transferability by training a network based on Fourier operators [23] using real data from a specific biological tissue type and testing it on different types. Despite this advancement, such approaches may struggle to generate high-quality images when confronted with test samples that significantly differ from the training data or originate from entirely different distributions. The work of Chen et al. [4] is closely related to ours, as it also performs spatial super-resolution. However, while the exact methodology employed by [4] to perform such task remains undisclosed, several key distinctions set our work apart: (i) We explicitly incorporate image alignment and registration to leverage aliased information for spatial super-resolution. (ii) Our network architecture prioritizes interpretability and is closely related to [5], enhancing robustness to new unseen real-world data. (iii) Our network operates independently of changes in illumination wavelength and/or detector/object distance, eliminating the need for retraining when these variables are altered. Rivenson et al. [31] introduced a supervised technique where an input hologram is first back-propagated through free space to generate an initial estimate of the latent field which is then fed into a multi-scale CNN that refines it. Huang et al. [19] proposed a self-supervised learning approach where a network similar to that of [3] is trained using a physics consistency loss with a strategy similar to that of [7] where the loss is evaluated between the input hologram and a simulated one using the predicted complex field distribution.

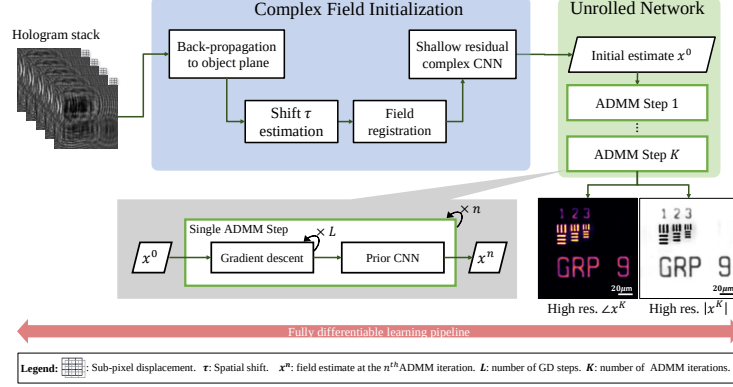


Fig. 2: The overall fully differentiable architecture of HoloADMM: A stack of low resolution noisy holograms captured at different heights used to reconstruct a high-quality and spatially super-resolved complex field.

3 Methodology

In this section the image formation model for in-line holography will be presented along with the problem formulation and our proposed model.

3.1 Image Formation Model

The schematics in Fig. 1 illustrate the basic setup of Gabor’s in-line lensless holographic imaging system which we use in our DIHM. Given a latent complex transmission field at the object plane $\mathbf{x} \in \mathbb{C}^{hw}$ sampled at high resolution with spatial dimensions $h \times w$, for each height z_i , $i = 1, \dots, N$ a hologram can be simulated using the following equation:

$$f_{s,\tau_i,z_i}(\mathbf{x}) = D_s W_{\tau_i} |P_{z_i} \mathbf{x}|^2 \quad (1)$$

Where $f_{s,\tau,\mathbf{z}} : \mathbb{C}^{hw} \mapsto \mathbb{R}^{h'w'}$ is the forward in-line holographic image formation model. The latent field $\mathbf{x}$ is propagated to the detector plane using the complex near-field Fresnel propagation kernel P_{z_i} [22]. The real valued hologram is obtained by calculating the modulus square of $P_{z_i} \mathbf{x}$ which is then spatially warped using the matrix W_τ simulating spatial shifts in the (u, v) plane (see Fig. 1) with sub-pixel accuracy, defined by the set $\tau = \{(\tau_1^u, \tau_1^v), \dots, (\tau_N^u, \tau_N^v)\}$. These spatial shifts are necessary for enabling image super-resolution capability, see Sec. 4.3. D_s is a down-sampling matrix that reduces the image size by a factor of s , resulting in dimensions $h' = h/s$ and $w' = w/s$. Notice that the image formation model as described here is non-linear, more general, and physically accurate where the object of interest is assumed to have both absorption and phase shift properties—though absorption can sometimes be negligible for thin and transparent micro-organisms. Additionally, sensor read and shot noise

sources are also simulated using the noise model from [11]: $\sigma(k) = \sqrt{\alpha \cdot y(k) + \gamma}$, where σ is the pixel-dependent standard deviation of the noise level at pixel k , with α and γ representing the variances of shot and read noises, respectively, and $y(k)$ is the clean input pixel value.

3.2 Problem Formulation

The aim of this work is to reconstruct a spatially super-resolved complex transmission distribution from a given stack of noisy low-resolution input holograms $H = [\mathbf{h}_1, \dots, \mathbf{h}_N]$ captured at N different heights $\mathbf{z} = [z_1, \dots, z_N]$ with spatial shifts $\tau = \{(\tau_1^u, \tau_1^v), \dots, (\tau_N^u, \tau_N^v)\}$. This problem can be reformulated as a regularized least squares minimization:

$$(\tilde{\mathbf{x}}, \tilde{\tau}) = \arg \min_{\mathbf{x}, \tau} \frac{1}{2} \|H - f_{s, \tau, \mathbf{z}}(\mathbf{x})\|_2^2 + \beta \Psi(\mathbf{x}) \quad (2)$$

Where Ψ is a regularizer that constrains the possible solution space on the distribution of the latent field $\mathbf{x}$ and β is a discrepancy parameter controlling the strength of such regularization. Accurate prediction of $\mathbf{z}$ is important in order to reconstruct $\mathbf{x}$ as the field at the detector plane can be propagated to the object plane and vice-versa using $P_{\mathbf{z}}$ or its conjugate $P_{\mathbf{z}}^*$. The value of $\mathbf{z}$ can be accurately estimated using any computational refocusing technique from the literature [35]. Spatial shifts τ need to be estimated with sub-pixel accuracy for multi-frame image registration, thereby enabling spatial image super-resolution. The optimization framework as expressed in Eq. 2 is non-linear in all optimization variables $(\tilde{\mathbf{x}}, \tilde{\tau})$ and can be solved by iteratively optimizing for one variable at a time while keeping the other one fixed, i.e., by alternating the following (A) and (B) steps:

$$(A) \quad \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|H - f_{s, \tau, \mathbf{z}}(\mathbf{x})\|_2^2 + \beta \Psi(\mathbf{x}) \quad (3)$$

$$(B) \quad \tilde{\tau} = \arg \min_{\tau} \frac{1}{2} \|H - f_{s, \tau, \mathbf{z}}(\mathbf{x})\|_2^2 \quad (4)$$

3.3 HoloADMM: End-to-end Learning For QPI

Solving for the latent field (A): Eq. 3 is a regularized non-linear least squares with no closed-form solution for $\tilde{\mathbf{x}}$: a good approximation can be obtained iteratively with a proper image prior Ψ such as Total Variation [33] or other natural image priors. Choosing a proper Ψ to solve Eq. 3 is not trivial: depending on the target scene properties multiple possible image priors can be used. We propose to learn a suitable prior agnostic to scene properties owing to the large representation capacity of deep neural networks. To this end, a variable splitting technique is used to solve Eq. 3 namely the Alternating Direction Method of Multipliers (ADMM) [12] where an auxiliary variable $\mathbf{v}$ is introduced such that:

$$(\tilde{\mathbf{x}}, \tilde{\mathbf{v}}) = \arg \min_{\mathbf{x}, \mathbf{v}} \frac{1}{2} \|H - f_{s, \tau, \mathbf{z}}(\mathbf{x})\|_2^2 + \beta \Psi(\mathbf{v}) \quad s.t. \quad \mathbf{x} - \mathbf{v} = 0 \quad (5)$$

Note that Eq. 5 is now a constrained version of the previous formulation in Eq. 3 where the data-fidelity and prior terms are no longer coupled and consequently they can be solved separately, Eq. 5 can be further split into multiple sub-problems by first retrieving its scaled augmented Lagrangian:

$$\mathcal{L}_\rho^{\text{ADMM}}(\mathbf{x}, \mathbf{v}, \mathbf{u}) = \frac{1}{2} \|H - f_{s,\tau,\mathbf{z}}(\mathbf{x})\|_2^2 + \beta\Psi(\mathbf{v}) + \frac{\rho}{2} \|\mathbf{x} + \mathbf{u} - \mathbf{v}\|_2^2 + \frac{\rho}{2} \|\mathbf{u}\|_2^2 \quad (6)$$

where $\mathbf{u}$ is the scaled Lagrange multiplier and ρ is a penalty term for the constraint in Eq. 5 forcing the final estimate $\mathbf{v}$ to be as close as possible to the true solution $\mathbf{x}$. To minimize Eq. 5 the saddle point of Eq. 6 needs to be found by iteratively solving the following three sub-problems:

$$\mathbf{x} \leftarrow \arg \min_{\mathbf{x}} \mathcal{L}_\rho^{\text{ADMM}}(\mathbf{x}, \mathbf{v}, \mathbf{u}) = \arg \min_{\mathbf{x}} \frac{1}{2} \|H - f_{s,\tau,\mathbf{z}}(\mathbf{x})\|_2^2 + \frac{\rho}{2} \|\mathbf{x} + \mathbf{u} - \mathbf{v}\|_2^2 \quad (7)$$

$$\mathbf{v} \leftarrow \arg \min_{\mathbf{v}} \mathcal{L}_\rho^{\text{ADMM}}(\mathbf{x}, \mathbf{v}, \mathbf{u}) = \arg \min_{\mathbf{v}} \frac{\rho}{2} \|\mathbf{x} + \mathbf{u} - \mathbf{v}\|_2^2 + \beta\Psi(\mathbf{v}) \quad (8)$$

$$\mathbf{u} \leftarrow \mathbf{u} + \mathbf{x} - \mathbf{v} \quad (9)$$

The full model architecture is shown in Fig. 2: in each ADMM iteration (shown in the grey box), $\mathbf{x}$ in Eq. 7 is updated using multiple steps of plain gradient descent or any other gradient based approach, e.g., conjugate gradient. Since $\mathbf{x} = \mathbf{a} + i\mathbf{b} \in \mathbb{C}$, gradients are calculated using Wirtinger derivatives [26] where $\mathbf{x}$ is updated via:

$$\mathbf{x} \leftarrow \mathbf{x} - \alpha \cdot \frac{\partial}{\partial \mathbf{x}^*} \mathcal{L}_\rho^{\text{ADMM}}(\mathbf{x}, \mathbf{v}, \mathbf{u}) \quad (10)$$

Where $\frac{\partial}{\partial \mathbf{x}^*} = \frac{1}{2}(\frac{\partial}{\partial \mathbf{a}} + i\frac{\partial}{\partial \mathbf{b}})$, $\mathbf{x}^*$ is the complex conjugate of $\mathbf{x}$, and α is a learning rate. The gradient calculation requires defining the backward model of $f_{s,\tau,\mathbf{z}}$ which we denote with $f_{s,\tau,\mathbf{z}}^b$: it is approximated by up-sampling all N low resolution input holograms, back-propagating each hologram in the stack using $P_{z_i}^*$, $i = 1, \dots, N$ back to the object plane, and aligning the resulting complex images. Finally, the final back-propagated output $\mathbf{x}^b$ is obtained by averaging all N aligned complex fields. Eq. 8 can be viewed as a simple denoising problem with the identity matrix I as the forward model and a noisy input $x + u$, where the target is to estimate a clean complex field v . In principle, any plug-and-play denoiser should be suitable to solve Eq. 8. However, to achieve better performance and enable end-to-end learning we use a trainable Convolutional Neural Network (CNN) as a denoiser that acts as a learned image prior which in our case is a ResUnet architecture [40]. Finally, the update step of $\mathbf{u}$ is straightforward.

Fig. 3 shows a detailed implementation of the update rule for any given $\mathbf{x}$ where multiple steps of gradient descent are first performed to solve Eq. 7 followed by a denoising step to solve Eq. 8. Notice that the weights of the CNN are shared among all ADMM iterations which are unrolled to form the overall network architecture. The weights are learned in a supervised end-to-end manner together with all the other hyper-parameters namely the scaled Lagrange multiplier ρ from Eq. 6 and the gradient descent learning rate α from Eq. 10.

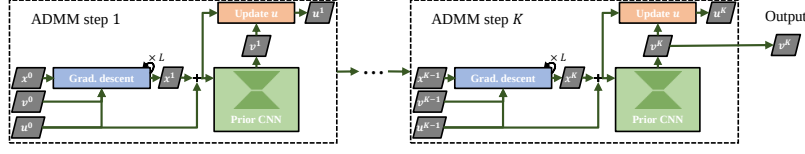


Fig. 3: Detailed implementation of each unrolled ADMM iteration.

Solving for spatial shifts (B): In principle any image registration algorithm can be used to estimate spatial shifts $\tau = \{(\tau_1^u, \tau_1^v), \dots, (\tau_N^u, \tau_N^v)\}$ between a reference frame (hologram closest to the object plane) and all the other frames. Note that the registration is not performed on the raw input holograms because diffraction patterns are different due to field propagation since the distance between the object and detector planes changes with each capture. The registration is carried out on the images of the back-propagated stack of N holograms. We use a fast FFT based alignment approach with arbitrary sub-pixel accuracy [17].

Complex shallow network: Since the target domain is complex by nature, the initial estimate ($\mathbf{x}^0 \in \mathbb{C}$) undergoes further processing through a shallow complex CNN [16] in order to learn close correlations between its real and imaginary components without the need to separate them into two distinct channels. Such network is designed with a residual connection linking the input and output distributions (see the suppl. mat. for more details), enhancing information flow and reducing noise in $\mathbf{x}^0$: it has multiple convolution layers with complex kernels in $\mathbb{C}$ thus the learned weights are complex in nature.

Initialization: As depicted in the initialization block in Fig. 2 (blue box), after estimating the relative spatial displacements using the back-propagated fields, $\mathbf{x}^0$ is obtained by applying the backward model $f_{s,\tau,\mathbf{z}}^b$ as described before and feeding the resulting complex field to the shallow residual complex network. $\mathbf{u}^0$ and $\mathbf{v}^0$ are set to 0. The initial value for the learning rate α of the plain gradient descent step in Eq. 10 is set to 0.01 and ρ in Eq. 6 is set to 0.1, recall that both of these variables are learned end-to-end (refer to the suppl. mat.). We use $L = 3$ gradient descent steps in each block for a total of $n = 5$ ADMM steps.

3.4 Datasets and Training Details

The lack of real holographic datasets, along with ground truth complex fields, is the main limitation hindering the development of learning-based reconstruction models. Self-supervised approaches, [19, 27] where the loss is evaluated between a real captured hologram and a simulated one using the predicted phase and amplitude distributions of the latent complex field, suffer from three major drawbacks: (i) the problem formulation is severely ill-posed since multiple possible solutions might correspond to the same measurement, eventually leading

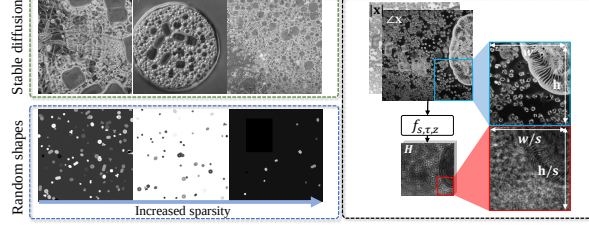


Fig. 4: Synthetic data generated using a stable diffusion model (top left) and a software that generate random shapes with different sparsity levels (bottom left). An input latent field and its simulated hologram (right).

to inaccurate reconstruction that does not necessarily correspond to the true target; (ii) designing effective self-supervised loss functions can be challenging. These loss function needs to capture relevant characteristics of the data and the reconstruction task, which may not always be straightforward to define; (iii) the real forward model, if not carefully designed, will further contribute to low reconstruction quality even with small simulation inaccuracies. Large scale real data with ground truth complex distributions is not trivial to collect, in fact, ground truth data used in the literature is obtained by the MH-PR algorithm [10] or other closely related variants which imposes an upper-bound hindering the true capability of any network. We therefore argue that learning an accurate inverse model through large-scale pixel-accurate synthetic data is crucial for high-quality reconstruction. To this end, we generate a large number of synthetic microscopic data featuring images with varying complexity and sparsity. A stable diffusion model [32] fine-tuned on microscopic images⁴ is used to generate dense interconnected samples with high spatial frequencies, in addition, samples with a varying degree of sparsity were obtained using a simple software that generates a random number of simple shapes. Fig. 4 shows some samples generated using the two modalities; sparse as well as dense data samples are generated to account for the real nature of microscopic images where individual or few cells as well as dense connective tissues might be present in a given image. Phase and amplitude distributions are obtained from a single gray-scale image $\mathbf{I}$ by $A = |\mathbf{x}| = e^{\omega \times \mathbf{I}}$, $\Phi = \angle \mathbf{x} = \pi \mathbf{I}$, where $\omega \leq 0$ is a weight determining the degree of transparency of the sample, no amplitude information (or fully transparent image) corresponds to $\omega = 0$, data is simulated with $\omega = -1.6$ to favor highly transparent samples. The dataset contains more than 100k different training samples (a smaller dataset led to lower performances). As shown on Fig. 4 (right) input low resolution noisy holograms are coupled with high-resolution clean phase and amplitude images. During training, wavelengths are chosen randomly from [440, 530, 638] nm and used to simulate each hologram along with a broad refocusing range from 0.5 mm up to 1.0 mm with a step size of $10\mu\text{m}$. The dataset is challenging because a single object can have many corresponding

⁴ https://huggingface.co/Fictiverse/Stable_Diffusion_Microscopic_model

Table 1: Quantitative comparison results on D_i and D_o synthetic test data without spatial super-resolution. (†) is a self-supervised approach.

Method	RMSE $\downarrow$ ($\times 10^{-3}$)				PSNR $\uparrow$				SSIM $\uparrow$			
	A		Φ		A		Φ		A		Φ	
	D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o
Wu et al. [37]	-	-	20.31	12.47	-	-	17.160	19.163	-	-	0.364	0.640
Huang et al. † [19]	-	-	7.81	12.50	-	-	14.271	17.715	-	-	0.602	0.549
Ren et al. $\times 1$ [30]	-	-	8.11	4.31	-	-	21.343	23.822	-	-	0.647	0.700
Rivenson et al. [31]	5.34	6.77	9.43	9.83	23.113	23.418	20.628	21.987	0.736	0.766	0.681	0.703
Chen et al. [3]	2.23	9.70	4.33	15.29	27.318	24.397	24.451	21.402	0.853	0.782	0.820	0.672
HoloADMM $\times 1$ (Ours)	1.23	0.96	2.42	1.55	29.373	30.198	26.410	28.092	0.923	0.905	0.902	0.877

holographic measurements each with different illumination wavelength and/or refocusing distance, forcing any learning-based approach to effectively learn the inverse model which is agnostic to changes in the input. Sub-pixel shifts are randomly simulated in the range of ± 3 pixels. HoloADMM is trained using the MSE as loss for 100 epochs with the Adam optimizer and a learning rate of $1e^{-4}$.

4 Results and Discussions

Results on synthetic as well as real in-line holographic data are presented in this section along with quantitative and qualitative comparisons with competing approaches. Additional qualitative results can be found in the suppl. mat.

4.1 Synthetic Holographic Data

HoloADMM and its competitors are trained exclusively on synthetic data, as detailed in Sec. 3.4, and evaluated on both inner and outer synthetic test sets denoted as D_i and D_o respectively: the former comprises synthetic images generated using the approach outlined in Sec. 3.4, while the latter contains a handful of classic test targets taken from the Set14 dataset [18] (chosen image indices are [1, 2, 9, 10, 12]). HoloADMM is trained using $N = 10$ holograms, yet it is able to infer complex fields from an arbitrary number of input holograms, provided enough memory. As demonstrated below, our model consistently exhibits superior reconstruction quality compared to the state-of-the-art, even when provided with just a single hologram as input or when the input is heavily down-sampled. Quantitative metrics reported in Tab. 1 demonstrate HoloADMM’s efficacy on both datasets, outperforming competing approaches by considerable margins according to all reported metrics with an average PSNR improvement of over 4dB on the phase (Φ) compared to the second best approach on D_o . Performances are confirmed by Fig. 5 where visual inspection reveals HoloADMM’s capability to preserve image details while effectively suppressing sensor noise, resulting in clean and sharp phase images in contrast to competing methods (note that [3] fails to suppress signal-dependent noise and [31, 37] suffer from blur artifacts).

When considering also joint spatial super-resolution, HoloADMM is trained using a stack of holograms decimated by a factor of $\times 4$ with an input shape

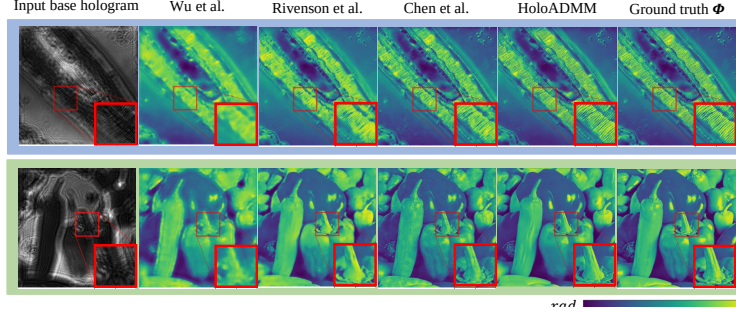


Fig. 5: Reconstructed Φ from some synthetic holographic samples from D_i and D_o .

Table 2: Quantitative comparison results on D_i and D_o synthetic test data with spatial super-resolution.

Method	SR factor	RMSE $\downarrow$ ($\times 10^{-3}$)				PSNR $\uparrow$				SSIM $\uparrow$			
		D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o	D_i	D_o
Ren et al. [30]	$\times 4$	-	-	23.76	75.49	-	-	16.531	14.159	-	-	0.284	0.520
HoloADMM (Ours)	$\times 4$	2.46	1.32	4.59	1.94	26.544	29.313	23.764	27.626	0.806	0.827	0.769	0.791
HoloADMM (Ours)	$\times 8$	7.13	8.53	12.35	14.94	21.910	22.938	19.442	20.794	0.554	0.663	0.488	0.598

of 128×128 pixels, yet it achieves superior phase image quality (see, Fig. 6), compared to other competitors trained using high-resolution inputs (512×512 pixels). Tab. 2 quantitatively corroborates these results: HoloADMM reaches an SSIM value of 0.791 with $\times 4$ down-sampled inputs, compared to 0.703 obtained by [31] with full-resolution input holograms. Ren et al. [30] also performs spatial super-resolution via sub-pixel convolutions [34]. The results in Table 2 and Figure 6 demonstrate that our model produces sharper images, preserving high-frequency details and outperforming [30] for $\times 4$ super-resolution. Even with an extreme factor of $\times 8$, HoloADMM is capable of yielding reasonable results, (see Tab. 2), in contrast to [30], which fails to generate meaningful distributions (see suppl. mat. for more qualitative results). Additionally, results reported in Table 3 and Figure 7 compare the reconstruction performance of HoloADMM, already trained on synthetic data, with other model-based iterative methods on *Airplane*, *Barbara*, and *Baboon* samples from D_o . For a fair comparison, we report results using single as well as 10 holograms as input; in both cases, our approach outperforms iterative solvers and effectively suppresses signal-dependent sensor noise, unlike [7, 38], where the proposed algorithms tend to fit the noise model as the number of iterations increases. Furthermore, HoloADMM preserves small spatial details, such as the fine whiskers in Figure 7, while avoiding undesirable artifacts like those produced by [27] in an attempt to suppress sensor noise. Iterative approaches typically require a large amount of time to produce reasonable results (up to 15 min. for a 512×512 image for [7]), while HoloADMM, once trained, can infer complex field distributions in under a second.

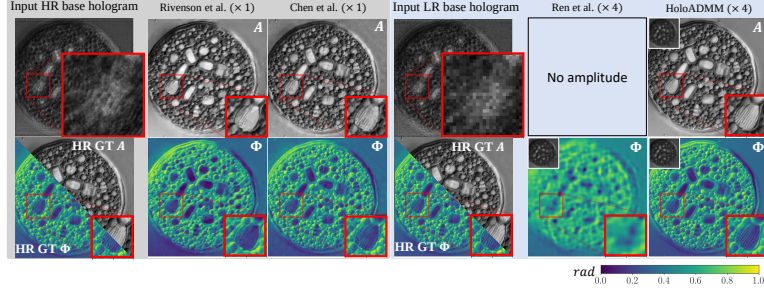


Fig. 6: Reconstructions with $\times 4$ SR for HoloADMM and [30] and $\times 1$ for [3, 31].

Table 3: Quantitative results (PSNR) on some standard test targets. The table compares the PSNR achieved by our approach with other iterative methods. Computation times refer to an NVIDIA A6000 except for (*) that use the CPU only.

Method	# holograms	Time (s)	Airplane		Barbara		Baboon	
			A	Φ	A	Φ	A	Φ
Niknam et al. [27]	1	529	21.019	11.020	16.127	14.261	17.425	13.583
Chen et al. [7]	1	990	18.297	17.4709	14.720	12.051	15.403	14.960
Chen et al. [38]	1	797	17.188	18.012	20.047	17.179	15.504	14.744
SH-PR [10]	1	60*	16.307	9.085	17.540	12.937	14.934	12.753
MH-PR [10]	10	100*	19.465	11.295	19.900	19.565	20.596	18.567
HoloADMM $\times 1$	1	0.29	<u>29.178</u>	<u>24.828</u>	<u>25.261</u>	<u>23.959</u>	<u>23.800</u>	<u>21.454</u>
HoloADMM $\times 1$	10	<u>0.58</u>	32.223	28.205	28.289	26.757	27.198	25.020

4.2 Real Holographic Data

Fig. 1 shows our DIHM prototype with multi-height phase retrieval and sub-pixel shift super-resolution capability. Please refer to suppl. mat. for details on the hardware setup. We capture holograms of two different samples: **Polystyrene beads**, which mimics real microscopic samples in both structural size and light transmission properties. They have diameters ranging from $0.9\text{-}9\text{ }\mu\text{m}$ and are equipped with carboxyl functional groups on their surface, allowing them to covalently attach to the cover glass and form a monolayer structure. The cover glass slides are attached to a sticky microchannel slide (IBIDI sticky-Slide I Luer). The sample is imaged using an illumination wavelength in the visible range, e.g., 638 nm, to achieve ambiguity-free phase retrieval, prevent phase wrapping, and provide suitable phase contrast. The channel is filled with a tailored liquid solution (Immersion). **A phase calibration target “Phasefocus”**, was used to calibrate our DIHM setup. This target is fabricated through Reactive Ion Etching (RIE) of amorphous SiO_2 , which is patterned via optical lithography, as described in [15]. The target includes both phase and amplitude features, which can be imaged using our DIHM. The phase features are 600 nm deep trenches etched into transparent amorphous SiO_2 , providing feature sizes that span a wide range of spatial frequencies, from length scales of $2\text{ }\mu\text{m}$ to length scales of $600\text{ }\mu\text{m}$. Our focus is on the reconstruction performance of the smallest features, such as GRP 9,

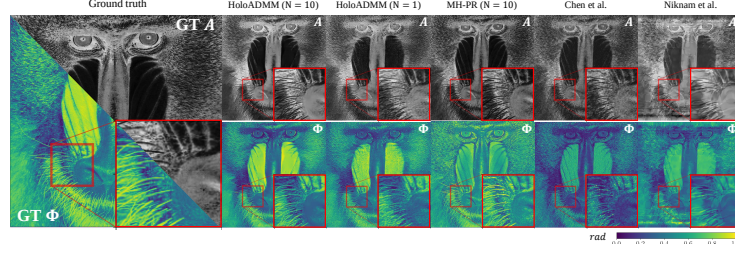


Fig. 7: Reconstruction results of A and Φ on Baboon test target.

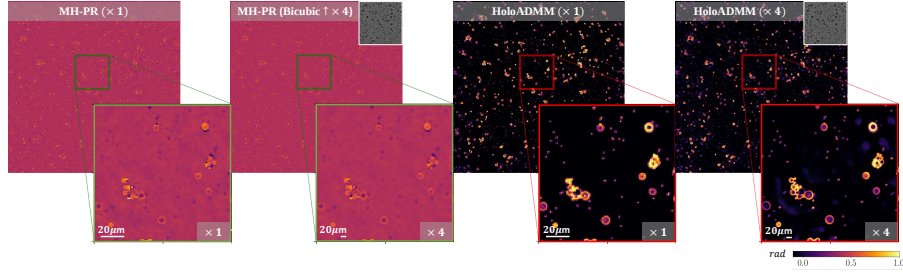


Fig. 8: Reconstructed Φ from real beads holograms with $\times 1$ and $\times 4$ SR. Results from MH-PR are also shown with $\times 4$ resolution using bicubic up-sampling. Input low resolution holograms are shown on the top right corners.

which have a structure size of 2-10 μm . We tested all competing models listed in Tab. 1 trained solely on synthetic data and obtained very poor and sometimes meaningless reconstructions on real data, further highlighting the challenge of model transferability beyond the training domain. Qualitative results for those models are reported in the suppl. mat. In contrast, HoloADMM is able to produce high-quality reconstructions on different holographic samples for a model trained on synthetic data only and without any form of adaptation. Notably, it not only produces meaningful results but also achieves higher spatial resolution, up to $\times 4$ as shown in Fig. 8 reaching an effective pixel size of 280 nm, alongside results from the standard MH-PR algorithm. Fig. 9 shows reconstructed GRP 9 features from the phase calibration target, along with outputs from other iterative approaches listed in Tab. 3. HoloADMM reconstructs cleaner and sharper phase and amplitude images suppressing noise and undesirable diffraction patterns, revealing fully transparent features with higher contrast and allowing for better quality 3D surface reconstructions from predicted Φ w.r.t. competitors.

4.3 Ablation Studies

Different ablation experiments have been conducted on a small subset of the training dataset. The increase in the number of ADMM steps, i.e., the number of unrolled blocks, leads to lower overall loss values, as shown in Fig. 10 (left),

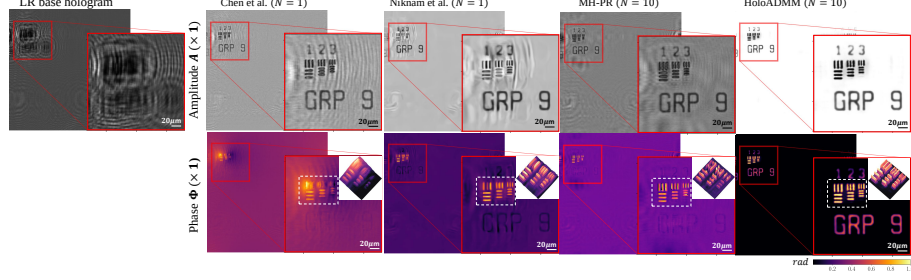


Fig. 9: Reconstructed A and Φ from a real phase calibration target. Highlighted features are invisible in the bright-field domain but exhibit high phase contrast. Reconstructed 3D surfaces from Φ of the etched lines are shown on the top right corner of each zoomed-in region.

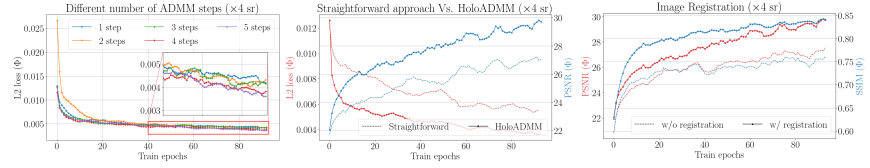


Fig. 10: Ablation experiments: (*left*) with different number of unrolled ADMM steps, (*center*) with a straightforward approach, (*right*) without image registration.

but requires more computational resources; we opted for $n = 5$ in our model. A naive approach that takes a hologram stack as input and predicts the latent field falls short of achieving good performance, even on training data, as shown in Fig. 10 (center), and is unable to generalize beyond that domain. Image registration is crucial for image super-resolution; without such a step, complementary spatial information in the form of aliasing is not exploited, and the spatial quality of the reconstructed images deteriorates, as shown in Fig. 10 (right). More ablation results are in the suppl. mat.

5 Conclusions

We introduced HoloADMM, an approach that combines interpretability of model-based solvers and the large learning capacity of deep neural networks. We demonstrated that by leveraging large-scale synthetic datasets, high-quality phase imaging capability can be achieved. Additionally, our approach exhibits strong generalization abilities, seamlessly extending to real captured holographic data with notable accuracy. This not only highlights the promise of our methodology but also suggests its potential for practical applications across diverse domains, from biomedical imaging to materials science and beyond. Nonetheless, there are other avenues for potential investigation, such as joint computational refocusing. This is a natural extension of our work and will be further investigated.

Acknowledgments

This collaborative work was funded by Sony Group Corporation’s Research and Development Center. It was also partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”) and by the Italian Ministry for Education, University and Research (MIUR) under the Departments of Excellence Initiative, project “Internet of Things: methods, technology and applications”.

References

1. Brady, D.J., Choi, K., Marks, D.L., Horisaki, R., Lim, S.: Compressive holography. *Optics express* **17**(15), 13040–13049 (2009)
2. Burch, C., Stock, J.: Phase-contrast microscopy. *Journal of Scientific Instruments* **19**(5), 71 (1942)
3. Chen, H., Huang, L., Liu, T., Ozcan, A.: Fourier imager network (fin): A deep neural network for hologram reconstruction with superior external generalization. *Light: Science & Applications* **11**(1), 254 (2022)
4. Chen, H., Huang, L., Liu, T., Ozcan, A.: efin: Enhanced fourier imager network for generalizable autofocusing and pixel super-resolution in holographic imaging. *IEEE Journal of Selected Topics in Quantum Electronics* **29**(4: Biophotonics), 1–10 (2023)
5. Chen, N., Wang, C., Heidrich, W.: Holographic 3d particle imaging with model-based deep network. *IEEE Transactions on Computational Imaging* **7**, 288–296 (2021)
6. Chen, N., Wang, C., Heidrich, W.: Differentiable holography (June 2022). <https://doi.org/10.1002/lpor.202200828>
7. Chen, X., Wang, H., Razi, A., Kozicki, M., Mann, C.: Dh-gan: a physics-driven untrained generative adversarial network for holographic imaging. *Optics Express* **31**(6), 10114–10135 (2023)
8. Cintora, P., Arikath, J., Kandel, M., Popescu, G., Best-Popescu, C.: Cell density modulates intracellular mass transport in neural networks. *Cytometry Part A* **91**(5), 503–509 (2017)
9. Denis, L., Lorenz, D., Thiébaud, E., Fournier, C., Trede, D.: Inline hologram reconstruction with sparsity constraints. *Optics letters* **34**(22), 3475–3477 (2009)
10. Fienup, J.R.: Reconstruction of an object from the modulus of its fourier transform. *Optics letters* **3**(1), 27–29 (1978)
11. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing* **17**(10), 1737–1754 (2008)
12. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications* **2**(1), 17–40 (1976)
13. Gabor, D.: Microscopy by reconstructed wave-fronts. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **197**(1051), 454–487 (1949)

14. Gerchberg, R., Saxton, W.: A practical algorithm for the determination of phase from image and diffraction plane pictures. SPIE milestone series MS **93**, 306–306 (1994)
15. Godden, T.M., Muniz-Piniella, A., Claverley, J.D., Yacoot, A., Humphry, M.J.: Phase calibration target for quantitative phase imaging with ptychography. *Opt. Express* **24**(7), 7679–7692 (2016)
16. Guberman, N.: On complex valued convolutional neural networks. arXiv preprint arXiv:1602.09046 (2016)
17. Guizar-Sicairos, M., Thurman, S.T., Fienup, J.R.: Efficient subpixel image registration algorithms. *Optics letters* **33**(2), 156–158 (2008)
18. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5197–5206 (2015)
19. Huang, L., Chen, H., Liu, T., Ozcan, A.: Self-supervised learning of hologram reconstruction using physics consistency. *Nature Machine Intelligence* **5**(8), 895–907 (2023)
20. Kemper, B., Von Bally, G.: Digital holographic microscopy for live cell applications and technical inspection. *Applied optics* **47**(4), A52–A61 (2008)
21. Latychevskaia, T., Fink, H.W.: Solution to the twin image problem in holography. *Physical review letters* **98**(23), 233901 (2007)
22. Latychevskaia, T., Fink, H.W.: Practical algorithms for simulation and reconstruction of digital in-line holograms. *Applied optics* **54**(9), 2424–2434 (2015)
23. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895 (2020)
24. Mir, M., Bhaduri, B., Wang, R., Zhu, R., Popescu, G.: Quantitative phase imaging. *Progress in optics* **57**(133–37), 217 (2012)
25. Momey, F., Denis, L., Olivier, T., Fournier, C.: From fienup’s phase retrieval techniques to regularized inversion for in-line holography: tutorial. *JOSA A* **36**(12), D62–D80 (2019)
26. Mshimba, A.S.A., Tutschke, W.: Functional analytic methods in complex analysis and applications to partial differential equations. World Scientific (1995)
27. Niknam, F., Qazvini, H., Latifi, H.: Holographic optical field recovery using a regularized untrained deep decoder network. *Scientific reports* **11**(1), 10903 (2021)
28. Park, Y., Depeursinge, C., Popescu, G.: Quantitative phase imaging in biomedicine. *Nature photonics* **12**(10), 578–589 (2018)
29. Park, Y., Diez-Silva, M., Popescu, G., Lykotrafitis, G., Choi, W., Feld, M.S., Suresh, S.: Refractive index maps and membrane dynamics of human red blood cells parasitized by plasmodium falciparum. *Proceedings of the National Academy of Sciences* **105**(37), 13730–13735 (2008)
30. Ren, Z., Xu, Z., Lam, E.Y.: End-to-end deep learning framework for digital holographic reconstruction. *Advanced Photonics* **1**(1), 016004–016004 (2019)
31. Rivenson, Y., Zhang, Y., Günaydin, H., Teng, D., Ozcan, A.: Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications* **7**(2), 17141–17141 (2018)
32. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
33. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* **60**(1–4), 259–268 (1992)

34. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016)
35. Tamamitsu, M., Zhang, Y., Wang, H., Wu, Y., Ozcan, A.: Comparison of gini index and tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical wavefront. arXiv preprint arXiv:1708.08055 (2017)
36. Vest, C., Sweeney, D.: Holographic interferometry of transparent objects with illumination derived from phase gratings. *Applied Optics* **9**(10), 2321–2325 (1970)
37. Wu, Y., Wu, J., Jin, S., Cao, L., Jin, G.: Dense-u-net: dense encoder–decoder network for holographic imaging of 3d particle fields. *Optics Communications* **493**, 126970 (2021)
38. Xiwen, C., Hao, W., Zhao, Z., Zhenmin, L., Huayu, L., Tong, Y., Abolfazl, R.: Enhancing digital hologram reconstruction using reverse-attention loss for untrained physics-driven deep learning models with uncertain distance. In: SPIE Photonics West 2024 (2024)
39. Zhang, W., Cao, L., Brady, D.J., Zhang, H., Cang, J., Zhang, H., Jin, G.: Twin-image-free holography: a compressive sensing approach. *Physical review letters* **121**(9), 093902 (2018)
40. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* **15**(5), 749–753 (2018)
41. Ziemczonok, M., Kuś, A., Wasylczyk, P., Kujawińska, M.: 3d-printed biological cell phantom for testing 3d quantitative phase imaging systems. *Scientific reports* **9**(1), 18872 (2019)