OpticalNet: An Optical Imaging Dataset and Benchmark Beyond the Diffraction Limit

Benquan Wang[1*] Ruyi An[4*] Jin-Kyu So[1] Sergei Kurdiumov[3] Eng Aik Chan[1]
Giorgio Adamo[1] Yuhan Peng[1] Yewen Li[1✉] Bo An[1,2]
[1]Nanyang Technological University [2]Skywork AI, Singapore [3]University of Southampton
[4]The University of Texas at Austin

https://Deep-See.github.io/OpticalNet

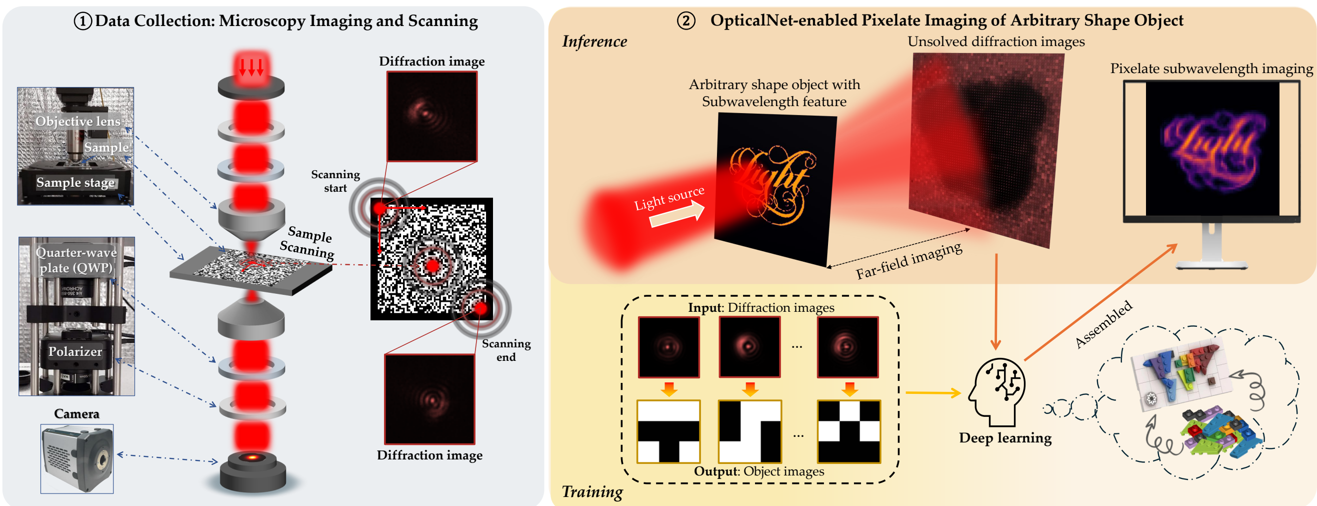


Figure 1. **Framework of OpticalNet.** Drawing an analogy to modular construction, where small units could be assembled to create larger complex objects, we build the OpticalNet dataset that deconstructs arbitrary-shaped objects into basic building blocks–small $n \times n$ grid regions consisting of squares with sizes below the diffraction limit. This dataset is collected through microscopy imaging via sample scanning, and we can train a deep-learning-based model to predict object images using diffraction images as inputs. With the trained model, we translate diffraction images of complex-shaped objects into their corresponding object images for each spatial position and assemble these modular predictions accordingly to reconstruct the complete structures, enabling subwavelength imaging beyond the diffraction limit.

Abstract

Optical imaging capable of resolving nanoscale features would revolutionize scientific research and engineering applications across biomedicine, smart manufacturing, and semiconductor quality control. However, due to the physical phenomenon of diffraction, the optical resolution is limited to approximately half the wavelength of light, which impedes the observation of subwavelength objects such as the native state coronavirus, typically smaller than 200 nm. Fortunately, deep learning methods have shown remarkable potential in uncovering underlying patterns within data, promising to overcome the diffraction limit by revealing the mapping pattern between diffraction images and their corresponding ground truth object images. However, the absence of suitable datasets has hindered progress in this field–collecting high-quality optical data of subwavelength objects is highly difficult as these objects are inherently invisible under conventional microscopy, making it impossible to perform standard visual calibration and drift correction. Therefore, we provide the first general optical imaging dataset based on the "building block" concept for challenging the diffraction limit. Drawing an analogy to modular construction principles, we construct a comprehensive optical imaging dataset comprising subwavelength fundamental elements, i.e., small square units that can be assembled into larger and more complex objects. We then frame the task as an image-to-image translation task and evaluate various vision methods. Experimental results validate our "building block" concept, demonstrating that models trained on basic square units can effectively generalize to realistic, more complex unseen objects. Most importantly, by highlighting this underexplored AI-for-science area and its potential, we aspire to advance optical science by fostering collaboration with the vision and machine learning communities.

*Equal contributions.
✉Corresponding to Yewen Li <*yewen001@e.ntu.edu.sg*>.

1. Introduction

The opportunity to glimpse the wonders of the tiny world with one's eyes has fascinated researchers for millennia. From rudimentary magnifying glasses of ancient times to today's advanced microscopes, this journey has given rise to the fields of optical imaging and microscopy [12, 62], which have become indispensable tools in fundamental research and engineering applications, such as biostructure imaging [5, 67, 71] and precision manufacturing [86, 119]. However, the wave nature of light manifests in diffraction [115], a universal phenomenon that becomes particularly pronounced when light waves interact with structures of dimensions comparable to the wavelength, fundamentally limiting the observation resolution in optical systems. An illustrative explanation is provided in Fig. 2. This limitation, known as the **diffraction limit** [87], constrains the minimum observable feature in the imaging plane to a subwavelength scale $d = \lambda/(2\mathrm{NA})$, where λ denotes the illumination wavelength and NA is the numerical aperture. Consequently, conventional optical microscopy using visible light is restricted to a spatial resolution of approximately 200 ~ 250 nm [31].

This constraint led to electron microscopy (EM) [24] development, which achieves atomic-scale resolution[48, 90] but requires complex sample preparation and vacuum environments [11, 73]. More critically, the irreversible radiation damage from high-energy electron beams prevents their application in real-time imaging of live biological entities in their native state such as the inspection of the SARS-CoV-2 virus [23, 104]. In contrast, optical microscopy enables non-invasive, real-time observation with simple sample preparation and prolonged observation capability [9, 92, 115], although its resolution is fundamentally constrained by the diffraction limit. To overcome this limitation while preserving optical advantages, various optical super-resolution techniques have been developed. Notably, super-resolution fluorescence microscopy [36, 84], recognized by the 2014 Nobel Prize in Chemistry [8], achieved resolution of tens of nanometers. However, this approach requires invasive fluorescence tagging and complex sample preparation [7, 17], compromising the inherent benefits of optical imaging and limiting its application in real-time imaging and semiconductor metrology [26, 69]. This prompts a fundamental question: "*Can we see objects beyond the diffraction limit with only conventional microscopy?*"

Fortunately, deep learning methods have shown remarkable potential in uncovering the underlying patterns within data [50]. In addition, the ability of neural networks to efficiently solve the inverse scattering problem has also been demonstrated [93], providing a solid theoretical foundation for using deep learning[30, 99]. Therefore, this insight enables us to resolve optical imaging at subwavelength resolution in an end-to-end image-to-image translation manner [39, 100, 118]. The interaction of light with objects creates

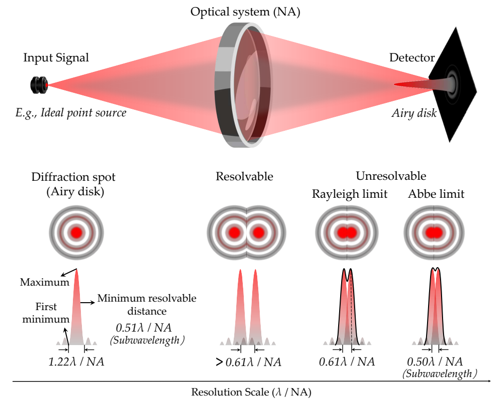


Figure 2. Illustration of the **diffraction limit**. Similar to how digital images cannot have infinitely small pixels, an ideal point light source inevitably diffracts into a finite-sized Airy disk in the imaging plane. Then, two adjacent diffraction spots become indistinguishable when their separation falls below a certain threshold.

diffraction images that contain transformed detailed metrological information about the objects being observed, such as shape, size, and position [15, 78, 95]. Such metrological information of an object being observed could be represented by a 2D image, termed *object image*. Given the diffraction images, neural networks can be directly utilized to decode them to the ground truth object image. With the help of vision algorithms, such an end-to-end approach requires no sample modification or tagging, operates at low light intensities to avoid photobleaching and does not rely on non-linear light-matter interactions [3]. This presents a distinct advantage over existing optical methods for overcoming the diffraction limit, even owning the potential to achieve molecular and atomic-scale resolution of live biological entities [56].

Deep learning methods require extensive training data. However, to the best of our knowledge, there exists no open-source subwavelength imaging dataset serving the general purpose of addressing fundamental diffraction limit challenges. While several microscopic image datasets do exist, they are highly domain-specific and constrained to particular imaging targets, such as observing biological cells [70, 107] and conducting lithic use-wear analysis [110] at scales larger than the subwavelength. Additionally, the diversity in optical setups, data formats, and experimental configurations among these datasets prevents them from being collectively used to train models that can generalize to observing different and complex objects. This lack of a high-quality, generalizable dataset significantly hinders the advancement of optical imaging beyond the diffraction limit. Therefore, there is an urgent need for an open-source microscopic dataset at the subwavelength scale that can be widely used by the vision and machine learning communities.

To provide a dataset that could be generalizable, we adopt a building blocks approach where fundamental subwavelength square elements can be assembled into arbitrary complex shapes.

Our contributions are summarized as follows:

- In collaboration with top optical scientists, we provide the first optical building blocks concept imaging dataset beyond the diffraction limit. This required extremely careful design and precise execution using advanced Focused Ion Beam (FIB) technology at nanometer-scale precision, alongside a high-precision custom-built microscopy system with sophisticated stabilization methods. Given the significant costs involved, we also provide simulation code for proof-of-concept testing before conducting actual experiments.
- To evaluate the generalization ability of the trained model, we provide two special testing sets with deeply subwavelength-scale features: *i)* "Light" testing set for evaluating the performance in observing objects with arbitrary shapes; and *ii)* "Siemens Star" (SS) testing set for evaluation on arbitrary rotations and arbitrary size.
- For algorithm benchmarking, we formulate the problem as an image-to-image translation task, specifically pixel-level binary classification. Through evaluating a wide range of vision methods, we gain important insights for future research—notably, transformers focusing on global information outperform CNN-based methods in handling environmental noise. Experimental results demonstrate the feasibility of our concept, enabling the possibility of overcoming the diffraction limit with traditional optical microscopy.

By open-sourcing this optical imaging dataset and benchmark, we seek to encourage interdisciplinary collaboration between optical science and computer vision communities to address current challenges in subwavelength optical imaging. This dataset provides a foundation for exploring computational approaches that enhance conventional microscopy's capabilities beyond the diffraction limit. Such advancements could potentially benefit a wide range of applications where high-resolution imaging is critical, including biological specimen analysis such as virus screening and industrial applications like semiconductor quality control.

2. Related Work

Optical Methods to Challenge the Diffraction Limit. Traditional optical methods like scanning near-field optical microscopy [44] offer high resolution but require invasive near-field probes and cannot image internal structures. Fluorescence-based methods [34, 52] achieve nanometer resolution but require invasive fluorescent labeling. Ptychography [14, 63, 83, 85] represents a promising alternative achieves subwavelength resolution but faces challenges including long acquisition times, computational intensity for phase retrieval algorithms. These limitations have spurred interest in AI-enhanced solutions[3, 77, 80, 97]. Recent advances show that AI-enabled methods can achieve deep subwavelength resolution through non-invasive far-field measurements without complex post-processing [57, 76, 96, 101, 102], demonstrating a promising direction in optical research.

Image-to-image Translation. Image-to-image translation [39, 100, 118] is a core computer vision task aimed to learn mappings between input and output images, facilitating tasks like image segmentation [58, 64], style transfer [45, 61, 88, 109, 118], image colorization [19, 38, 41, 49, 51, 112], and image restoration [74, 103, 111, 113]. Outstanding performance has been achieved on common objects, such as medical segmentation using U-Net [82] through end-to-end training. However, most approaches rely on the premise that the correspondence between inputs and outputs can be visually discernible, for instance, segmenting pixels into categories or transferring visual styles without changing objects' structure. However, in optical research, such direct visual correspondence is not always observable with traditional microscopy. To bridge this gap and complement existing tasks, we introduce a new vision challenge: translating diffracted images to clear object images at the subwavelength level.

Microscopic Image Datasets. In the realm of vision tasks involving microscopic images, numerous applications span various scientific disciplines [1, 2, 6, 10, 18, 20, 25, 28, 37, 42, 43, 47, 53, 54, 65, 66, 75, 105, 108, 114, 116]. Representative studies include research on bacteria [98], biological cells [13, 16, 40], tissue types [94], and material structures [32, 110]. Each application presents unique challenges, particularly in terms of the high-level detail and precision required in the images. These challenges are often compounded by issues such as ambiguity in object properties and variations in sensing modalities. To address these challenges, our approach includes a versatile framework that supports both fundamental atomic objects and practical objects across simulated and realistic modalities. A key distinction to existing datasets is the provision of an easy-to-use simulation procedure for generating synthetic samples with diverse object properties and sensing modalities, considering the prohibitive cost of creating new image samples. This approach allows researchers to economically validate new ideas via simulation, before the costly experimental sample acquisition, thereby conserving resources and human effort.

3. OpticalNet Dataset

Our primary contribution is the provision of a comprehensive optical imaging dataset, that combines the theoretical simulation data for systematic exploration and experimental data for real-world verification. Termed the OpticalNet, the dataset comprises fundamental square unit samples that can be assembled to form objects of arbitrary and complex shapes. This foundational dataset contains image-to-image translation relationships between elementary square objects and their corresponding diffraction images, providing the basis for training neural networks to translate diffraction images back to their corresponding central elementary geometric structures, namely object images. We provide a dataset datasheet [27] in Appendix A.

3.1. Data Acquisition

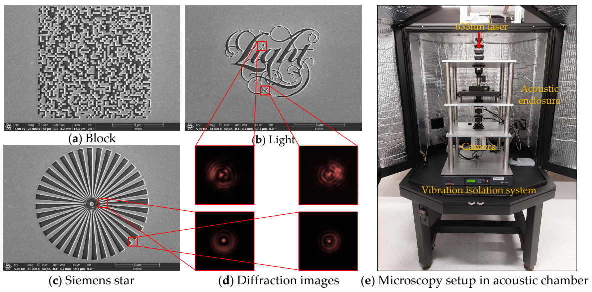


Figure 3. **a~d:** Fabricated samples and the diffraction images; **e:** High-precision microscopy for data acquisition.

Given the large size of object samples and the limited reception field of a camera, we adopt a practical approach of photographing only sub-regions of an object sample at a time. This method leverages the knowledge that any complex object forming an object image can be systematically decomposed through spatial scanning into smaller sub-regions, each producing its own diffraction image.

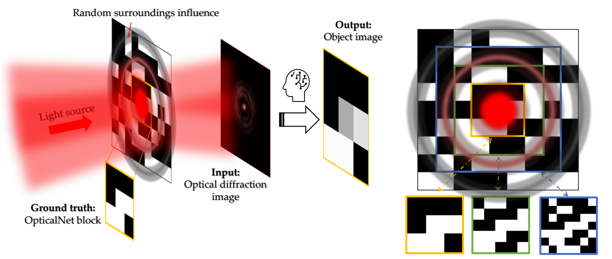
Sample Fabrication. We fabricate our samples using a high-precision dual-beam FIB system [79], a nanofabrication technique extensively used across various sectors, including semiconductor manufacturing and quantum computing. This system employs a focused ion beam to precisely mill material, complemented by an integrated electron microscope that enables monitoring of the fabrication process. The samples are prepared on a 130 nm Au film on a glass coverslip, a configuration commonly used in optical and electronic devices. The fabrication process produces three types of representative test samples: the Block sample, consisting of squares that are uniform in size (180 nm) and positioned randomly without overlapping; a calligraphic "Light" sample demonstrating complex curved features and geometries; and a Siemens star—a benchmark for testing optical resolution [35], characterized by a radial pattern of periodic straight lines. These samples are shown in Fig. 3.

High-Quality Microscopic Imaging. To obtain high-quality subwavelength optical imaging data for each sub-region sample, we employ an ultra-precision custom-built microscopy system and carefully design a set of stabilization methods to enable long-term stable high-precision imaging with minimal mechanical drift and environment noise, as depicted in Fig. 3. This system utilizes a coherent light source of 633nm wavelength in a vertically aligned configuration to achieving an effective pixel size of 41.7 nm on the sample plane. A linearly polarized beam of light is then focused onto the sample plane through a high-numerical-aperture objective mounted on a precision piezoelectric stage. The detection scheme adopts a symmetric configuration, where the transmitted intensity diffraction signals are detected with a high-sensitivity sCMOS camera positioned in the far-field regime. The optical path is mechanically stabilized through a commercial vibration isolation system, and the acoustic chamber enclosing the whole optical setup significantly attenuates environmental noise across the acoustic frequency range. We leave more details of the optical imaging process, such as the positioning accuracy method of the optical imaging system, in Appendix C. With extensive high-precision and stabilization measures, as well as operation by professional optical scientists, we have ensured the quality of the collected dataset to the best of our efforts.

Simulation Framework for Proof of Concept. As the realistic data creation process is slow and highly costly, we present an open-source computational framework for simulating optical field propagation using the angular spectrum method [60]. This framework provides simulation of light-matter interactions and diffraction phenomena. Developed entirely in Python, this framework combines computational efficiency (~ 1 second per instance on an AMD EPYC 7742) with user-friendliness, enabling researchers to easily modify and extend the codebase for various applications. Researchers can fine-tune illumination characteristics, *e.g.*, wavelength, and customize sample properties to match specific experimental conditions. Using our OpticalNet dataset as a benchmark, the framework allows researchers to assess how variations in the physical size and shapes of samples influence model performance and generalization capabilities, by simply inputting binary mask images of the desired structures. The framework then generates the corresponding diffraction images, which researchers can use alongside object images to train neural networks and evaluate their performance on the provided samples, serving as a proof of concept before conducting realistic experiments.

3.2. Data Characteristics and Analysis

Categories of the Datasets. We categorize the dataset into three groups: Block dataset, Light dataset, and Siemens Star (SS) dataset. Detailed information can be found in Table 1. The diffraction images are sized 64×64 while the object images are provided in three different scales: 3×3, 5×5, and 7×7. The Block dataset serves as the training set, where object images of large scale can be built up from the square units within this dataset. The latter two are used to assess the generalization of the trained model. The "Light" dataset features complex curved boundaries and irregular structural elements with multiple scale ranges of sub-wavelength features that exceed the resolution limits of traditional optical microscopy. This dataset is useful for determining if the model has genuinely learned the underlying physics of diffraction rather than merely memorizing specific geometric images from the Block dataset. The SS dataset features a 36-spoke Siemens star pattern, serving as a standardized benchmark to



(a) OpticalNet block with random surroundings (b) OpticalNet of 3×3 / 5 × 5 / 7 × 7 block

Figure 4. **(a)** An optical Block with random surroundings demonstrates how the diffraction image is influenced by square units outside the target 3×3 region. **(b)** Different sizes of optical Blocks are used to investigate how varying sizes of ground truth images affect the model's ability.

Table 1. Dataset categories, usage, and the number of data points for training and testing.

Dataset		Block	Light	SS
For training?		✔	✗	✗
Simulation	# training samples	12,068	-	-
	# testing samples	1,000	4,356	4,356
Experiment	# training samples	26,316	-	-
	# testing samples	4,356	4,356	4,356

test the optical imaging resolution ability of models across continuous size variations and arbitrary angular orientations.

As illustrated in Fig. 4, while the incident light field exhibits maximum intensity within a 3×3 central block area defined by its Full Width at Half Maximum (FWHM), a diffraction image is not simply determined by this area but is a complex interference including contributions from the surroundings. This fact makes the task significantly challenging, prompting us to explore the optimal size of the ground truth object images. Therefore, a key design of the dataset is the provision of 3 sizes of ground truth images, including 3×3, 5×5, and 7×7. In the 3×3 case, the trained model must focus on the central part while contending with the influence of the surroundings. Conversely, in the 7×7 case, the model may receive insufficient information to accurately predict all square units due to the diminished light intensity on the outer edges. According to the number of white squares, the statistics of realistic experiment data are shown in Fig. 5. The statistics of simulation data are presented in Appendix C.1. Note that, the high number of all-black blocks and the low number of blocks containing a large number of white squares are due to the scanning process, explained in Appendix C.2.

Difficulty Level by the Dataset. The simpler scenario focuses on sub-wavelength OpticalNet Block data. By training on the Block dataset and testing on previously unseen testing samples of similar structures, we assess the model's ability to learn fundamental diffraction principles in elementary square units. On the other hand, the hard case is the model's generalization learned from the Block in uniform size and orientation to objects of arbitrary size and direction. This includes the SS and Light datasets. In particular, the SS dataset, featuring the Siemens Star presents the most significant challenge. The manufactured Siemens Star sample includes 36 spokes arranged uniformly across 360°, with the distance between them decreasing continuously from the outer edge to the center, testing the model's capability to resolve details at various scales and directions.

4. Task Definition: Image-to-Image Translation

With the collected data, we now define our task as an image-to-image translation problem characterized by a mapping function $\mathcal{F}_\phi : \mathbb{R}^{H_1 \times W_1 \times 1} \to \mathbb{R}^{H \times W \times 1}$ parameterized by a neural network ϕ. This function $\mathcal{F}$ is designed to transform input diffraction images into outputs that approximate the corresponding ground truth object images. Formally, our dataset comprises N diffraction images, denoted as $\{x_i \in \mathbb{R}^{H_1 \times W_1 \times 1}\}_{i=1}^N$, and their respective ground truth object images denoted as $\{y_i \in \mathbb{R}^{H \times W \times 1}\}_{i=1}^N$. To optimize our image translation mapping function $\mathcal{F}$, we define a fundamental loss function for training:

$$\mathcal{L}_\mathcal{F} = \frac{1}{N} \sum_{i=1}^{N} \ell(\mathcal{F}(x_i), y_i), \tag{1}$$

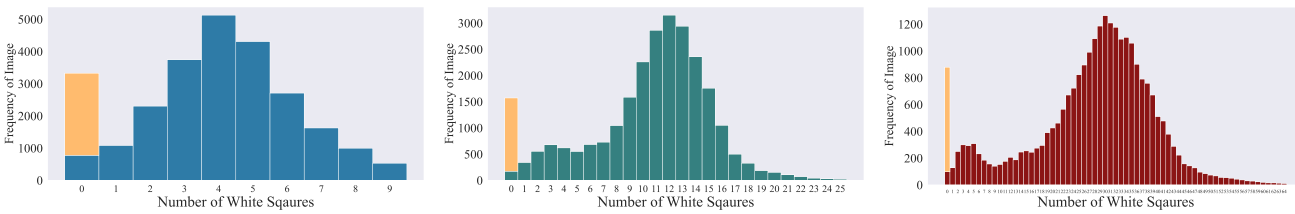
where $\ell(\cdot, \cdot)$ is a loss function quantifying the discrepancy between the model's predicted image $\mathcal{F}(x_i)$ and the corresponding ground truth image y_i.

As the ground truth object image y_i uses binary values to indicate object presence at each pixel, we employ the Binary Cross-Entropy (BCE) loss function:

$$\ell(\mathcal{F}(x_i), y_i) = -\frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} [y_i^{(h,w)} \log \mathcal{F}(x_i)^{(h,w)} + (1 - y_i^{(h,w)}) \log(1 - \mathcal{F}(x_i)^{(h,w)})], \tag{2}$$

where $\mathcal{F}(x_i)^{(h,w)}$ and $y_i^{(h,w)}$ denote the predicted and ground truth values at pixel location (h, w), respectively. The neural network parameterizing $\mathcal{F}$ could be optimized using gradient-based techniques, such as SGD [81] and Adam [46], to minimize $\mathcal{L}_\mathcal{F}$ in Eq. 1 over the training dataset. Once trained, this model is capable of predicting object images from previously unseen diffraction images. Though others may also use regression methods or even generative methods to model Eq. 1, to maintain simplicity in this study, we have particularly chosen to define the task of translating a diffraction image to a ground truth object image as a pixel-level binary classification problem and leave the exploration of other modeling methods in the future work.

Evaluation Using Stiching A simple procedure assessing the quality or visualizing the final output could utilize a threshold λ that converts the predicted probabilities from $\mathcal{F}$ into binary classification images directly. Each pixel in the output image is labeled as either occupied by the object (1) or not (0), based on the predicted probability relative to this



(a) 3 × 3 (b) 5 × 5 (c) 7 × 7

Figure 5. Distribution showing the count of white squares (etched regions to allow light transmission) within the experimental Block dataset. The orange bars represent additional all-black object images (unetched substrates) included in the dataset to help the model learn to discriminate environmental noise from actual object features.

threshold. The binarized output at each pixel location (h, w), represented by $\hat{y}_i^{(h,w)}$, is determined as follows:

$$\hat{y}_i^{(h,w)} = \begin{cases} 1 & \text{if } \mathcal{F}(x_i)^{(h,w)} \geq \lambda, \\ 0 & \text{if } \mathcal{F}(x_i)^{(h,w)} < \lambda, \end{cases} \tag{3}$$

The threshold λ is typically chosen based on validation performance or set to a default value such as 0.5.

Since our scanning process is moving at one square every step, a specific square unit's prediction could utilize all the block images that could cover it to enhance the prediction. As shown in Fig. 6, for a unit position (k, l) (marked as the red square), its output is related to multiple object images that can reach this position. For each object image of them, it could generate a corresponding diffraction image, and then we use a model taking this diffraction image as input, we would get a prediction image that contains the prediction for the location (k, l). Finally, we average the prediction on localization (k, l) using these block images. Therefore, we propose an enhanced evaluation and visualization procedure using a stitching process. Formally, this procedure could be expressed as

$$y_{sample}^{(k,l)} = \mathbb{E}_{x_m \sim \mathcal{X}'}[\mathcal{F}(x_m)^{(k',l')}], \tag{4}$$

where $\mathcal{X}'$ denotes a set of the diffraction images x_m whose corresponding object image could cover the localization (k, l) of a whole big sample, and (k', l') denotes the relevant position in the prediction image $\mathcal{F}(x_m)$. After this stitching process, we could then apply the same binary procedure in Eq. 3 to get the final binary classification result.

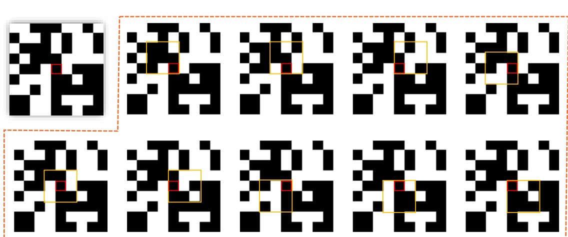
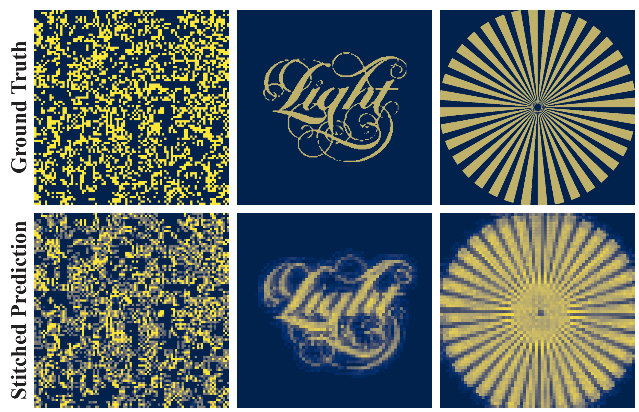


Figure 6. Illustration for the stitching. For the 3×3 block configuration setting, each target location (red box) is covered by nine overlapping block images (yellow box).

Table 2. Comparisons of models trained on simulation Block dataset evaluated on different test sets. Best result is marked in **bold**.

Method	Block			Light			SS		
	acc.	F1	JI	acc.	F1	JI	acc.	F1	JI
ResUNet-a	72.10	65.23	65.94	72.73	69.30	65.02	62.13	52.20	44.81
AttU-Net	75.34	67.21	67.59	73.34	69.18	65.10	61.65	54.77	47.30
ResNet-18	81.51	78.31	68.94	75.33	76.47	72.65	69.19	**63.99**	52.72
ResNet-34	83.48	79.44	69.73	**75.62**	76.86	72.12	66.12	60.28	51.94
transformer	**84.77**	**79.51**	**71.30**	75.00	**77.62**	**73.82**	**69.43**	62.79	**53.19**



(a) Block (b) Light (c) SS

Figure 7. Visualization of stitched predictions using the transformer model on the Block datasets.

5. Algorithm Benchmarking

Our experiments engage in comparisons with several state-of-the-art vision models on our OpticalNet dataset. Through the experiments, we aim to **1)** assess the feasibility of using optical patterns learned from the Block dataset to perform image-to-image translation tasks on more complex, unseen shapes that extend beyond merely small-dimensional squares, and hence the ability to construct meaningful, generalized patterns from the learning; **2)** evaluate the fidelity of simulation datasets in reflecting the trends observed in datasets collected from realistic experiment and hence evaluate the framework's practical applicability; **3)** for vision and machine learning communities, examine potential trends exhibited by the models to better understand their effectiveness and application in optimal pattern recognition.

Datasets. These models trained on the Block datasets are tested against specific Block test sets and generalized test sets of unseen, more complex object images.

Evaluation Metrics. To evaluate model performance on our

Table 3. Performance under metrics of models trained on experiment datasets with varying ground truth block dimensions., evaluated across different test sets. Best result for each configuration is marked in **bold**.

GT dimension	Method	Block			Light			SS		
		acc.	F1	JI	acc.	F1	JI	acc.	F1	JI
3×3	ResUNet-a	67.32	51.18	45.70	69.34	67.15	61.92	49.89	35.04	28.39
	AttU-Net	68.31	52.30	45.90	71.51	68.33	61.20	51.16	36.18	30.16
	ResNet-18	73.70	62.00	56.71	73.95	73.82	72.77	52.31	42.70	39.14
	ResNet-34	75.01	61.28	56.46	74.05	75.98	71.99	50.98	43.40	40.35
	transformer	**80.31**	**76.33**	**66.90**	**74.71**	**76.59**	**76.34**	**55.81**	**47.38**	**42.89**
5×5	ResUNet-a	66.17	54.89	45.61	73.47	67.04	60.32	53.82	36.61	30.16
	AttU-Net	68.92	53.21	44.94	73.32	67.89	59.45	51.15	37.90	32.13
	ResNet-18	73.19	60.35	56.99	74.30	74.90	69.56	50.96	43.07	37.60
	ResNet-34	74.09	60.94	57.51	74.99	75.51	72.55	51.53	42.16	38.05
	transformer	**80.17**	**74.36**	**63.51**	**77.98**	**78.35**	**77.49**	**53.50**	**48.37**	**41.84**
7×7	ResUNet-a	66.12	53.18	43.15	73.36	68.91	62.40	52.64	38.21	32.18
	AttU-Net	66.35	54.37	46.16	71.65	67.74	60.39	51.01	39.05	32.90
	ResNet-18	76.38	63.36	59.31	77.72	75.05	70.63	49.72	44.10	40.88
	ResNet-34	76.74	64.01	59.58	77.51	74.16	71.99	50.67	44.75	41.03
	transformer	**79.95**	**74.92**	**62.78**	**78.11**	**76.85**	**72.47**	**52.70**	**48.74**	**42.44**

image-to-image translation task, we employ classification accuracy, F1-score, and the Jaccard index (JI), averaged across the classes to assess the model by the ground truth.

Baseline Methods. We employ a variety of vision backbone methods, including ResUNet-a [22], Attention U-Net (AttU-Net) [68], ResNet-18, ResNet-34 [33], and transformer [91].

Implementation Details We train the models using a single NVIDIA A100 GPU and PyTorch [72], employing the Adam optimizer [46]. The initial learning rate is set at 1e-3, with a linear decay factor of 0.9 applied every 30 epochs. Training is conducted over a total of 500 epochs with a mini-batch size of 16. We set λ to the default value of 0.5 in Eq. 3. Data augmentation techniques that maintain the inherent characteristics of the optical images [29, 59], such as flips and right-angled rotations, are employed. Specifically, vertical and horizontal flips are each applied with an equal probability of occurrence. Rotations in increments of 90° are uniformly applied across 0, 90, 180, and 270°. **Detailed experimental setup are in Appendix D**.

5.1. Results on Simulation Dataset

We first train the model on the Block using the simulation dataset. As a proof of concept, we utilize a 3×3-grid setting of the Block configuration to validate the idea of building blocks under the simulation before expanding to datasets collected from experiments.

Table 2 reports the performance of various models on the simulated dataset. All models demonstrate strong performance on the Block patterns and the out-of-domain Light test set. The transformer achieves the best overall performance across all tests, while ResNet-18 performs better with the Light logo. Though performance dips on SS test set due to the models' difficulties with more subtle pattern variations

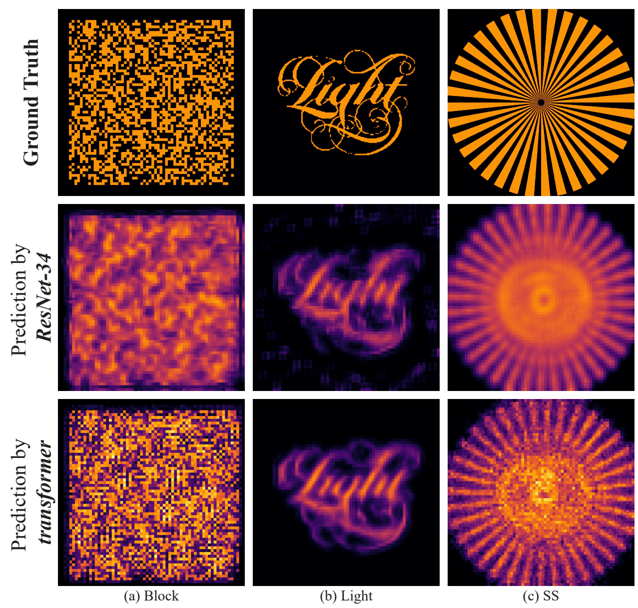
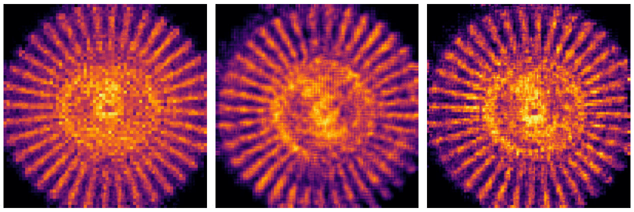


Figure 8. Visualization of stitched predictions using ResNet-34 (**row 2**) and transformer (**row 3**) on the experimental dataset. In (a), the transformer achieves a high-fidelity translation of the ground truth for the Block, whereas ResNet-34's output appears blurry. For (b) and (c) transformer resolves the spokes of SS with greater depth and preserves details of Light like small curves. Notably, ResNet-34 shows noisy predictions around the Light symbol.

in the central regions, the visualization of the composed Siemens Star pattern in Fig. 7, shows that the models are capable of effectively translating its broader components.

Given the encouraging results on the simulation dataset, it appears that our approach can effectively generalize to more complex and meaningful patterns. We now aim to extend this validation to the realistic experimental dataset.



(a) 3×3 (b) 5×5 (c) 7×7

Figure 9. Stitched predictions on SS performed by transformers trained with varying ground truth block dimensions.

5.2. Results on Realistic Experiment Dataset

Building on the findings from the simulated dataset, experiments presented in Table 3 assess the performance of models trained on realistic experimental datasets with configurations of 3×3, 5×5, and 7×7 units of the Block datasets respectively. From the quantitative result in Table 3 and the qualitative visualization depicted in Fig. 8, the models demonstrate the capability to learn the Block patterns, and generalize them to a broader variety of shapes, effectively translating these into unseen, more complex shape patterns. We observe minimal variation in results across the GT block dimensions used for training. The transformer consistently outperforms other models on both in-domain Block test sets and broader pattern recognition tasks. While the ResNet-based architectures previously matched the transformer's performance in simulations, they exhibit a decline and produce noisy output in the more complex experimental settings, which are inherently subject to greater environmental noise. This decline could be ascribed to the convolutional networks' focus on local information and their limited capacity for handling global information crucial in the optics domain for mitigating susceptibility to noise [4, 55]. Meanwhile, transformers, which process longer-term dependencies, may better manage the noise, potentially explaining their enhanced performance in realistically collected experimental datasets. Overall, there remains a high level of consistency in the trends observed from the simulation to the experiment dataset, validating the fidelity of the simulation and underscoring the robustness of our simulation-to-experiment modeling approach.

Additionally, we performed comparative analyses of stitched predictions using transformers trained on ground truth blocks of increasing sizes. As shown in Fig. 9, when the GT block dimensions increase, overall visual quality improves. However, this comes with a tradeoff of increased noise, particularly observed at the portions of spoke further from the center. This may be attributed to more blocks representing more information channels, but comes with additional noise susceptibility and computational complexity.

6. Impact and Limitation

Scientific and Engineering Impact. By integrating AI with microscopy, our work bridges machine learning with optical physics. Collaborations between two communities aim to enhance our understanding of subwavelength phenomena, deepening our insights into the underlying physics and broadening its applications. For instance, our dataset could enhance the resolution of viral particles with only a smartphone-based microscope [67, 89, 104], democratizing subwavelength imaging and potentially allowing for accurate remote infectious disease detection via on-chip microscopy [117]. Additionally, our tagging-free approach eliminates the need for harmful chemicals used in fluorescent microscopy, promoting the sustainability of microscopic imaging [106].

Limitation and Future Work. While our OpticalNet dataset showcases the capability of pixelated imaging beyond the diffraction limit, achieving continuous imaging remains both promising and challenging. Future work could explore finer resolution with more advanced computer vision algorithms. Additionally, this building block concept could be extended to 3D imaging through block stacking and RGB imaging using varying light wavelength, potentially enabling multi-dimensional, full-color super-resolution imaging. Moreover, our modest performance on the Siemens Star test (~ 50% accuracy) reveals room for improvement. While we successfully capture basic structural information, the predicted object images still have room for improvement in resolution and clarity. Future research could leverage advanced deep learning architectures and image processing algorithms to further enhance predicted image quality and resolution. Besides, the challenging sim-to-real task is also an interesting future work, which may improve cost efficiency by adapting a model trained on simulation data to realistic data.

7. Conclusion

We introduce a general optical imaging dataset beyond the diffraction limit and demonstrate that deep learning-based computer vision methods can effectively translate diffraction images into object images at subwavelength resolution. We showed that with the building block concept, models trained on fundamental square units can generalize to complex shapes. Our work offers a data-driven perspective while traditional approaches to challenge the diffraction limit have primarily focused on specialized optical concept. In our view based on the information theory, the deep learning training process incorporates prior knowledge that helps extract hidden subwavelength information from conventional microscopy data, enabling optical imaging capabilities beyond the diffraction limit. While our current implementation achieves promising results and has uncovered foundational insights, there remains ample room for exploration in future work. By fostering collaboration between the optical science and computer vision communities, we believe that the diffraction limit and imaging lens will be the things of the past, enabling new scientific discoveries and practical applications at the age of artificial intelligence and big data.

Acknowledgment

This work was supported by the Singapore National Research Foundation (Grant No. NRF-CRP23-2019-0006). The authors would like to thank Prof. Nikolay I. Zheludev for insightful discussion and inputs.

References

[1] Mary Damilola Aiyetigbo, Alexander Korte, Ethan Anderson, Reda Chalhoub, Peter Kalivas, Feng Luo, and Nianyi Li. Unsupervised microscopy video denoising. In *CVPR workshop*, 2024.

[2] Ashesh, Alexander Krull, Moises Di Sante, Francesco Silvio Pasqualini, and Florian Jug. μsplit: Image decomposition for fluorescence microscopy. In *ICCV*, 2023.

[3] Vasily N Astratov, Yair Ben Sahel, Yonina C Eldar, Luzhe Huang, Aydogan Ozcan, Nikolay Zheludev, Junxiang Zhao, Zachary Burns, Zhaowei Liu, Evgenii Narimanov, et al. Roadmap on label-free super-resolution imaging. *Laser & Photonics Rev.*, 17(12):2200029, 2023.

[4] Ravikiran Attota. Noise analysis for through-focus scanning optical microscopy. *Opt. Lett.*, 41(4):745–748, 2016.

[5] Harikrushnan Balasubramanian, Chad M Hobson, Teng-Leong Chew, and Jesse S Aaron. Imagining the future of optical microscopy: everything, everywhere, all at once. *Commun. Biol.*, 6(1):1096, 2023.

[6] Tal Ben-Haim and Tammy Riklin Raviv. Graph neural network for cell tracking in microscopy videos. In *ECCV*, 2022.

[7] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[8] Eric Betzig, Stefan W Hell, and William E Moerner. The nobel prize in chemistry 2014. *Nobel Media AB*, 2014.

[9] Martin J Booth. Adaptive optical microscopy: the ongoing quest for a perfect image. *Light Sci. Appl.*, 3(4):e165–e165, 2014.

[10] Nicolas Bourriez, Ihab Bendidi, Ethan Cohen, Gabriel Watkinson, Maxime Sanchez, Guillaume Bollot, and Auguste Genovesio. ChAda-ViT: Channel adaptive attention for joint representation learning of heterogeneous microscopy image. In *CVPR*, 2024.

[11] John J. Bozzola and Lonnie D. Russell. *Electron Microscopy: Principles and Techniques for Biologists*. Jones & Bartlett Learning, 1999.

[12] David J Brady. *Optical imaging and spectroscopy*. John Wiley & Sons, 2009.

[13] Jordão Bragantini, Merlin Lange, and Loïc Royer. Large-scale multi-hypotheses cell tracking using ultrametric contours maps. In *ECCV*, 2024.

[14] Eunju Cha, Chanseok Lee, Mooseok Jang, and Jong Chul Ye. DeepPhaseCut: Deep relaxation in phase for unsupervised fourier phase retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(12):9931–9943, 2022.

[15] Eng Aik Chan, Carolina Rendón-Barraza, Benquan Wang, Tanchao Pu, Jun-Yu Ou, Hongxin Wei, Giorgio Adamo, Bo An, and Nikolay I Zheludev. Counting and mapping of subwavelength nanoparticles from a single shot scattering pattern. *Nanophotonics*, 12(14):2807–2812, 2023.

[16] Claire Lifan Chen, Ata Mahjoubfar, Li-Chia Tai, Ian K Blaby, Allen Huang, Kayvan Reza Niazi, and Bahram Jalali. Deep learning in label-free cell classification. *Sci. Rep.*, 6 (1):21471, 2016.

[17] Long Chen, Xingye Chen, Xusan Yang, Chao He, Miaoyan Wang, Peng Xi, and Juntao Gao. Advances of super-resolution fluorescence polarization microscopy and its applications in life sciences. *Comput. Struct. Biotechnol. J.*, 18:2209–2216, 2020.

[18] Minghao Chen, Mukesh Bangalore Renuka, Lu Mi, Jeff Lichtman, Nir Shavit, and Yaron Meirovitch. Learning to correct sloppy annotations in electron microscopy volumes. In *CVPR workshop*, 2023.

[19] Xiaoyan Cong, Yue Wu, Qifeng Chen, and Chenyang Lei. Automatic controllable colorization via imagination. In *CVPR*, 2024.

[20] Colin L. V. Cooke, Fanjie Kong, Amey Chaware, Kevin C. Zhou, Kanghyun Kim, Rong Xu, D. Michael Ando, Samuel J. Yang, Pavan Chandra Konda, and Roarke Horstmeyer. Physics-enhanced machine learning for virtual fluorescence microscopy. In *ICCV*, 2021.

[21] Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. Flashattention: Fast and memory-efficient exact attention with io-awareness. In *NeurIPS*, 2022.

[22] Foivos I. Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.*, 162:94–114, 2020.

[23] RF Egerton, P Li, and M Malac. Radiation damage in the tem and sem. *Micron*, 35(6):399–409, 2004.

[24] Ray F. Egerton. *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM*. Springer, 2013.

[25] Benjamin Gallusser and Martin Weigert. TRACKASTRA: transformer-based cell tracking for live-cell microscopy. In *ECCV*, 2024.

[26] Mahipal Ganji, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker. Real-time imaging of dna loop extrusion by condensin. *Science*, 360(6384):102–105, 2018.

[27] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna M. Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Commun. ACM*, 64(12):86–92, 2021.

[28] Anna S. Goncharova, Alf Honigmann, Florian Jug, and Alexander Krull. Improving blind spot denoising for microscopy. In *ECCV*, 2020.

[29] Ander Gracia Moisés, Ignacio Vitoria Pascual, José Javier Imas González, and Carlos Ruiz Zamarreño. Data augmentation techniques for machine learning applied to optical spectroscopy datasets in agrifood applications: A comprehensive review. *Sensors*, 23(20):8562, 2023.

[30] Yu Guan, Tingting Fang, Diankun Zhang, and Congming Jin. Solving Fredholm integral equations using deep learning. *Int. J. Appl. Comput. Math.*, 8(2), 2022.

[31] Xiang Hao, Cuifang Kuang, Zhaotai Gu, Yifan Wang, Shuai Li, Yulong Ku, Yanghui Li, Jianhong Ge, and Xu Liu. From microscopy to nanoscopy via visible light. *Light Sci. Appl.*, 2(10):e108–e108, 2013.

[32] Eric Hayman, Barbara Caputo, Mario Fritz, and Jan-Olof Eklundh. On the significance of real-world conditions for material classification. In *ECCV*, 2004.

[33] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[34] Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.*, 19(11):780–782, 1994.

[35] Roarke Horstmeyer, Rainer Heintzmann, Gabriel Popescu, Laura Waller, and Changhuei Yang. Standardizing the resolution claims for coherent microscopy. *Nature Photonics*, 10(2):68–71, 2016.

[36] Bo Huang, Mark Bates, and Xiaowei Zhuang. Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.*, 78(1):993–1016, 2009.

[37] Mude Hui, Zihao Wei, Hongru Zhu, Fei Xia, and Yuyin Zhou. MicroDiffusion: Implicit representation-guided diffusion for 3D reconstruction from limited 2D microscopy projections. In *CVPR*, 2024.

[38] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans. Graph.*, 35(4):110:1–110:11, 2016.

[39] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

[40] Hartland W Jackson, Jana R Fischer, Vito RT Zanotelli, H Raza Ali, Robert Mechera, Savas D Soysal, Holger Moch, Simone Muenst, Zsuzsanna Varga, Walter P Weber, et al. The single-cell pathology landscape of breast cancer. *Nature*, 578(7796):615–620, 2020.

[41] Xiaoyang Kang, Tao Yang, Wenqi Ouyang, Peiran Ren, Lingzhi Li, and Xuansong Xie. Ddcolor: Towards photo-realistic image colorization via dual decoders. In *ICCV*, 2023.

[42] Bashir Kazimi, Karina Ruzaeva, and Stefan Sandfeld. Self-supervised learning with generative adversarial networks for electron microscopy. In *CVPR*, 2024.

[43] Rihuan Ke, Aurélie Bugeau, Nicolas Papadakis, Peter Schütz, and Carola-Bibiane Schönlieb. Learning to segment microscopy images with lazy labels. In *ECCV*, 2020.

[44] Fritz Keilmann and Rainer Hillenbrand. Near-field microscopy by elastic light scattering from a tip. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 362(1817):787–805, 2004.

[45] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *ICML*, 2017.

[46] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[47] Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, Dominique Beaini, Maciej Sypetkowski, Chi Vicky Cheng, Kristen Morse, Maureen Makes, Ben Mabey, and Berton Earnshaw. Masked autoencoders for microscopy are scalable learners of cellular biology. In *CVPR*, 2024.

[48] Abinash Kumar, Jonathon N Baker, Preston C Bowes, Matthew J Cabral, Shujun Zhang, Elizabeth C Dickey, Douglas L Irving, and James M LeBeau. Atomic-resolution electron microscopy of nanoscale local structure in lead-based relaxor ferroelectrics. *Nature Mater.*, 20(1):62–67, 2021.

[49] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Learning representations for automatic colorization. In *ECCV*, 2016.

[50] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[51] Junsoo Lee, Eungyeup Kim, Yunsung Lee, Dongjun Kim, Jaehyuk Chang, and Jaegul Choo. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence. In *CVPR*, 2020.

[52] Mickaël Lelek, Melina T Gyparaki, Gerti Beliu, Florian Schueder, Juliette Griffié, Suliana Manley, Ralf Jungmann, Markus Sauer, Melike Lakadamyali, and Christophe Zimmer. Single-molecule localization microscopy. *Nature Rev. Methods Primers*, 1(1):39, 2021.

[53] Rui Li, Mikhail Kudryashev, and Artur Yakimovich. Solving the inverse problem of microscopy deconvolution with a residual Beylkin-Coifman-Rokhlin neural network. In *ECCV*, 2024.

[54] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O'Donnell, Heng Huang, Mei Chen, and Weidong Cai. Unsupervised instance segmentation in microscopy images via panoptic domain adaptation and task re-weighting. In *CVPR*, 2020.

[55] Sheng Liu, Michael J Mlodzianoski, Zhenhua Hu, Yuan Ren, Kristi McElmurry, Daniel M Suter, and Fang Huang. scmos noise-correction algorithm for microscopy images. *Nature Methods*, 14(8):760–761, 2017.

[56] Tongjun Liu, Jun-Yu Ou, Eric Plum, Kevin F MacDonald, and Nikolay I Zheludev. Visualization of subatomic movements in nanostructures. *Nano Let.*, 21(18):7746–7752, 2021.

[57] Tongjun Liu, Cheng-Hung Chi, Jun-Yu Ou, Jie Xu, Eng Aik Chan, Kevin F MacDonald, and Nikolay I Zheludev. Picophotonic localization metrology beyond thermal fluctuations. *Nature Mater.*, 22(7):844–847, 2023.

[58] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.

[59] Boyuan Ma, Xiaoyan Wei, Chuni Liu, Xiaojuan Ban, Haiyou Huang, Hao Wang, Weihua Xue, Stephen Wu, Mingfei Gao, Qing Shen, et al. Data augmentation in microscopic images for material data mining. *NPJ Comput. Mater.*, 6(1):125, 2020.

[60] Kyoji Matsushima and Tomoyoshi Shimobaba. Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Opt. Express*, 17 (22):19662–19673, 2009.

[61] Youssef Alami Mejjati, Christian Richardt, James Tompkin, Darren Cosker, and Kwang In Kim. Unsupervised attention-guided image-to-image translation. In *NeurIPS*, 2018.

[62] Jerome Mertz. *Introduction to optical microscopy*. Cambridge University Press, 2019.

[63] Jianwei Miao. Computational microscopy with coherent diffractive imaging and ptychography. *Nature*, 637:281–295, 2025.

[64] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(7):3523–3542, 2022.

[65] Björn Möller, Zhengyang Li, Markus Etzkorn, and Tim Fingscheidt. Low-resolution-only microscopy super-resolution models generalizing to non-periodicities at atomic scale. In *CVPR workshop*, 2024.

[66] Youssef S. G. Nashed, Frédéric Poitevin, Harshit Gupta, Geoffrey Woollard, Michael Kagan, Chun Hong Yoon, and Daniel Ratner. CryoPoseNet: End-to-end simultaneous learning of single-particle orientation and 3D map reconstruction from cryo-electron microscopy data. In *ICCV workshop*, 2021.

[67] Vasilis Ntziachristos. Going deeper than microscopy: the optical imaging frontier in biology. *Nature Methods*, 7(8): 603–614, 2010.

[68] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. In *MIDL*, 2018.

[69] Ndubuisi G Orji, Mustafa Badaroglu, Bryan M Barnes, Carlos Beitia, Benjamin D Bunday, Umberto Celano, Regis J Kline, Mark Neisser, Yaw Obeng, and AE Vladar. Metrology for the next generation of semiconductor devices. *Nature Electron.*, 1(10):532–547, 2018.

[70] Wei Ouyang, Fynn Beuttenmueller, Estibaliz Gómez-de Mariscal, Constantin Pape, Tom Burke, Carlos Garcia-López-de Haro, Craig Russell, Lucía Moya-Sans, Cristina de-la Torre-Gutiérrez, Deborah Schmidt, Dominik Kutra, Maksim Novikov, Martin Weigert, Uwe Schmidt, Peter Bankhead, Guillaume Jacquemet, Daniel Sage, Ricardo Henriques, Arrate Muñoz-Barrutia, Emma Lundberg, Florian Jug, and Anna Kreshuk. Bioimage model zoo: a community-driven resource for accessible deep learning in bioimage analysis. *BioRxiv*, 2022.

[71] Vimal Prabhu Pandiyan, Aiden Maloney-Bertelli, James A Kuchenbecker, Kevin C Boyle, Tong Ling, Zhijie Charles Chen, B Hyle Park, Austin Roorda, Daniel Palanker, and Ramkumar Sabesan. The optoretinogram reveals the primary steps of phototransduction in the living human eye. *Sci. Adv.*, 6(37):eabc1124, 2020.

[72] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

[73] Christopher J Peddie, Christel Genoud, Anna Kreshuk, Kimberly Meechan, Kristina D Micheva, Kedar Narayan, Constantin Pape, Robert G Parton, Nicole L Schieber, Yannick Schwab, et al. Volume electron microscopy. *Nature Rev. Methods Primers*, 2(1):51, 2022.

[74] Mangal Prakash, Alexander Krull, and Florian Jug. Fully unsupervised diversity denoising with convolutional variational autoencoders. In *ICLR*, 2021.

[75] Valeriya Pronina, Filippos Kokkinos, Dmitry V. Dylov, and Stamatios Lefkimmiatis. Microscopy image restoration with deep wiener-kolmogorov filters. In *ECCV*, 2020.

[76] Tanchao Pu, Jun-Yu Ou, Vassili Savinov, Guanghui Yuan, Nikitas Papasimakis, and Nikolay I Zheludev. Unlabeled far-field deeply subwavelength topological microscopy (DSTM). *Adv. Sci.*, 8(1):2002886, 2021.

[77] Pengfei Qi, Zhengyuan Zhang, Xue Feng, Puxiang Lai, and Yuanjin Zheng. A symmetric forward-inverse reinforcement framework for image reconstruction through scattering media. *Opt. Laser Technol.*, 179:111222, 2024.

[78] Carolina Rendón-Barraza, Eng Aik Chan, Guanghui Yuan, Giorgio Adamo, Tanchao Pu, and Nikolay I Zheludev. Deeply sub-wavelength non-contact optical metrology of sub-wavelength objects. *APL Photonics*, 6(6), 2021.

[79] Steve Reyntjens and Robert Puers. A review of focused ion beam applications in microsystem technology. *J. Micromech. Microeng.*, 11(4):287, 2001.

[80] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017.

[81] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Stat.*, pages 400–407, 1951.

[82] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

[83] Atreyee Saha, Salman Siddique Khan, Sagar Sehrawat, Sanjana S. Prabhu, Shanti Bhattacharya, and Kaushik Mitra. LWGNet - learned wirtinger gradients for fourier ptychographic phase retrieval. In *ECCV*, 2022.

[84] Steffen J Sahl, Stefan W Hell, and Stefan Jakobs. Fluorescence nanoscopy in cell biology. *Nature Rev. Mol. Cell Biol.*, 18(11):685–701, 2017.

[85] Fahad Shamshad, Asif Hanif, Farwa Abbas, Muhammad Awais, and Ali Ahmed. Adaptive Ptych: Leveraging image adaptive generative priors for subsampled fourier ptychography. In *ICCV*, 2019.

[86] Yuki Shimizu, Liang-Chia Chen, Dae Wook Kim, Xiuguo Chen, Xinghui Li, and Hiraku Matsukuma. An insight into optical metrology in manufacturing. *Meas. Sci. Technol.*, 32 (4):042003, 2021.

[87] Ernst HK Stelzer. Beyond the diffraction limit? *Nature*, 417 (6891):806–807, 2002.

[88] Matteo Tomei, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Art2Real: Unfolding the reality of artworks via semantically-aware image-to-image translation. In *CVPR*, 2019.

[89] Derek Tseng, Onur Mudanyali, Cetin Oztoprak, Serhan O Isikman, Ikbal Sencan, Oguzhan Yaglidere, and Aydogan Ozcan. Lensfree microscopy on a cellphone. *Lab Chip*, 10 (14):1787–1792, 2010.

[90] Knut W Urban. Is science prepared for atomic-resolution electron microscopy? *Nature Mater.*, 8(4):260–262, 2009.

[91] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[92] Mary Grace M Velasco, Mengyang Zhang, Jacopo Antonello, Peng Yuan, Edward S Allgeyer, Dennis May, Ons M'Saad, Phylicia Kidd, Andrew ES Barentine, Valentina Greco, et al. 3D super-resolution deep-tissue imaging in living mice. *Optica*, 8(4):442–450, 2021.

[93] V. Vemuri and Gyu-Sang Jang. Inversion of fredholm integral equations of the first kind with fully connected neural networks. *J. Franklin Institute*, 329(2):241–257, 1992.

[94] Quoc Dang Vu, Simon Graham, Tahsin Kurc, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Jayashree Kalpathy-Cramer, Tianhao Zhao, et al. Methods for segmentation and classification of digital microscopy tissue images. *Front. in Bioeng. Biotechnol.*, 7:433738, 2019.

[95] Benquan Wang, Yewen Li, Eng Aik Chan, Giorgio Adamo, Bo An, Zexiang Shen, and Nikolay I Zheludev. Optical localization of nanoparticles in sub-rayleigh clusters. In *The European Conference on Lasers and Electro-Optics*, page ch_p_8. Optica Publishing Group, 2023.

[96] Benquan Wang, Ruyi An, Eng Aik Chan, Giorgio Adamo, Jin-Kyu So, Yewen Li, Zexiang Shen, Bo An, and Nikolay I Zheludev. Retrieving positions of closely packed subwavelength nanoparticles from their diffraction patterns. *Appl. Phys. Lett.*, 124(15), 2024.

[97] Hongda Wang, Yair Rivenson, Yiyin Jin, Zhensong Wei, Ronald Gao, Harun Günaydın, Laurent A Bentolila, Comert Kural, and Aydogan Ozcan. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods*, 16(1):103–110, 2019.

[98] Hongda Wang, Hatice Ceylan Koydemir, Yunzhe Qiu, Bijie Bai, Yibo Zhang, Yiyin Jin, Sabiha Tok, Enis Cagatay Yilmaz, Esin Gumustekin, Yair Rivenson, et al. Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Sci. Appl.*, 9(1):118, 2020.

[99] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[100] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional GANs. In *CVPR*, 2018.

[101] Yu Wang, Eng Aik Chan, Carolina Rendón-Barraza, Yijie Shen, Eric Plum, Kevin F MacDonald, Jun-Yu Ou, and Nikolay I Zheludev. 3d positional metrology of a virus-like nanoparticle with topologically structured light. *Appl. Phys. Lett.*, 124(22), 2024.

[102] Yu Wang, Eng Aik Chan, Carolina Rendón-Barraza, Yijie Shen, Eric Plum, and Jun-Yu Ou. 2D super-resolution metrology based on superoscillatory light. *Adv. Sci.*, 11(38): 2404607, 2024.

[103] Martin Weigert, Loïc Royer, Florian Jug, and Gene Myers. Isotropic reconstruction of 3D fluorescence microscopy images using convolutional neural networks. In *MICCAI*, 2017.

[104] Robert Witte, Vardan Andriasyan, Fanny Georgi, Artur Yakimovich, and Urs F Greber. Concepts in light microscopy of viruses. *Viruses*, 10(4):202, 2018.

[105] Steffen Wolf, Manan Lalit, Katie McDole, and Jan Funke. Unsupervised learning of object-centric embeddings for cell instance segmentation in microscopy images. In *ICCV*, 2023.

[106] Richard Wombacher and Virginia W Cornish. Chemical tags: applications in live cell fluorescence imaging. *J. Biophotonics*, 4(6):391–402, 2011.

[107] Heming Yao, Phil Hanslovsky, Jan-Christian Huetter, Burkhard Hoeckendorf, and David Richmond. Weakly supervised set-consistency learning improves morphological profiling of single-cell images. In *CVPR*, 2024.

[108] Enze Ye, Yuhang Wang, Hong Zhang, Yiqin Gao, Huan Wang, and He Sun. Recovering a molecule's 3D dynamics from liquid-phase electron microscopy movies. In *ICCV*, 2023.

[109] Zili Yi, Hao (Richard) Zhang, Ping Tan, and Minglun Gong. DualGAN: Unsupervised dual learning for image-to-image translation. In *ICCV*, 2017.

[110] Jing Zhang, Irving Fang, Hao Wu, Akshat Kaushik, Alice Rodriguez, Hanwen Zhao, Juexiao Zhang, Zhuo Zheng, Radu Iovita, and Chen Feng. LUWA dataset: Learning lithic use-wear analysis on microscopic images. In *CVPR*, 2024.

[111] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *CVPR*, 2017.

[112] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S. Lin, Tianhe Yu, and Alexei A. Efros. Real-time user-guided image colorization with learned deep priors. *ACM Trans. Graph.*, 36(4):119:1–119:11, 2017.

[113] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(7):2480–2495, 2021.

[114] Yuelin Zhang, Pengyu Zheng, Wanquan Yan, Chengyu Fang, and Shing Shin Cheng. A unified framework for microscopy defocus deblur with multi-pyramid transformer and contrastive learning. In *CVPR*, 2024.

[115] Nikolay I Zheludev. What diffraction limit? *Nature Mater.*, 7(6):420–422, 2008.

[116] Ruofan Zhou, Majed El Helou, Daniel Sage, Thierry Laroche, Arne Seitz, and Sabine Süsstrunk. W2S: microscopy data with joint denoising and super-resolution for widefield to SIM mapping. In *ECCV*, 2020.

[117] Hongying Zhu, Serhan O Isikman, Onur Mudanyali, Alon Greenbaum, and Aydogan Ozcan. Optical imaging techniques for point-of-care diagnostics. *Lab Chip*, 13(1):51–67, 2013.

[118] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.

[119] Jinlong Zhu, Jiamin Liu, Tianlai Xu, Shuai Yuan, Zexu Zhang, Hao Jiang, Honggang Gu, Renjie Zhou, and Shiyuan Liu. Optical wafer defect inspection at the 10 nm technology node and beyond. *Int. J. Extreme Manuf.*, 4(3):032001, 2022.