Tahir et al. *Light: Science & Applications* (2022)11:42
https://doi.org/10.1038/s41377-022-00730-x

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

Adaptive 3D descattering with a dynamic synthesis network

Waleed Tahir[1], Hao Wang[1] and Lei Tian[1,2]✉

Abstract

Deep learning has been broadly applied to imaging in scattering applications. A common framework is to train a descattering network for image recovery by removing scattering artifacts. To achieve the best results on a broad spectrum of scattering conditions, individual "expert" networks need to be trained for each condition. However, the expert's performance sharply degrades when the testing condition differs from the training. An alternative brute-force approach is to train a "generalist" network using data from diverse scattering conditions. It generally requires a larger network to encapsulate the diversity in the data and a sufficiently large training set to avoid overfitting. Here, we propose an *adaptive learning* framework, termed dynamic synthesis network (DSN), which *dynamically* adjusts the model weights and *adapts* to different scattering conditions. The adaptability is achieved by a novel "mixture of experts" architecture that enables dynamically synthesizing a network by blending multiple experts using a gating network. We demonstrate the DSN in holographic 3D particle imaging for a variety of scattering conditions. We show in simulation that our DSN provides generalization across a *continuum* of scattering conditions. In addition, we show that by training the DSN entirely on simulated data, the network can generalize to experiments and achieve robust 3D descattering. We expect the same concept can find many other applications, such as denoising and imaging in scattering media. Broadly, our dynamic synthesis framework opens up a new paradigm for designing highly *adaptive* deep learning and computational imaging techniques.

Introduction

Deep learning (DL) has become a powerful technique for tackling complex yet important computational imaging problems[1], such as phase imaging[2–5], tomography[6–9], ghost imaging[10–13], lightfield microscopy[14,15], super-resolution imaging[16–18], enhancing digital holography[19–22], and imaging through scattering media[23–26]. Within these computational imaging applications, one of the prevalent problems is "descattering", or removing scattering artifacts. For this purpose, a deep neural network (DNN) is generally trained to perform descattering, either directly on the measurement[2–4,10,13,19,20,23,25] or on the object-space projection[11,12]. Alternative to this "end-to-end" framework, another approach is to employ a pretrained DNN as the learned prior in an iterative model-based reconstruction algorithm to progressively mitigate scattering artifacts[7,8].

While increasingly effective, the existing descattering DL frameworks are fundamentally impeded by an outstanding challenge. They generally demonstrate optimal performance only when the scattering condition in the testing data match well with the training data, and the performance degrades sharply when the scattering conditions are mismatched[23,25]. Thus, if a task requires working with many different cases of scattering, it generally needs to train multiple "expert" networks, each optimized for a specific scattering condition[9,13,27]. Such expert networks are very limiting since they require a priori knowledge of the scattering condition for optimally training the DNN. An alternative "brute-force" method is to train a single "generalist" network using a larger data set combining diverse scattering conditions[23,25]. However, this generalist approach often achieves lower performance

Correspondence: Lei Tian (leitian@bu.edu)
[1]Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA
[2]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
These authors contributed equally: Waleed Tahir, Hao Wang

© The Author(s) 2022
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

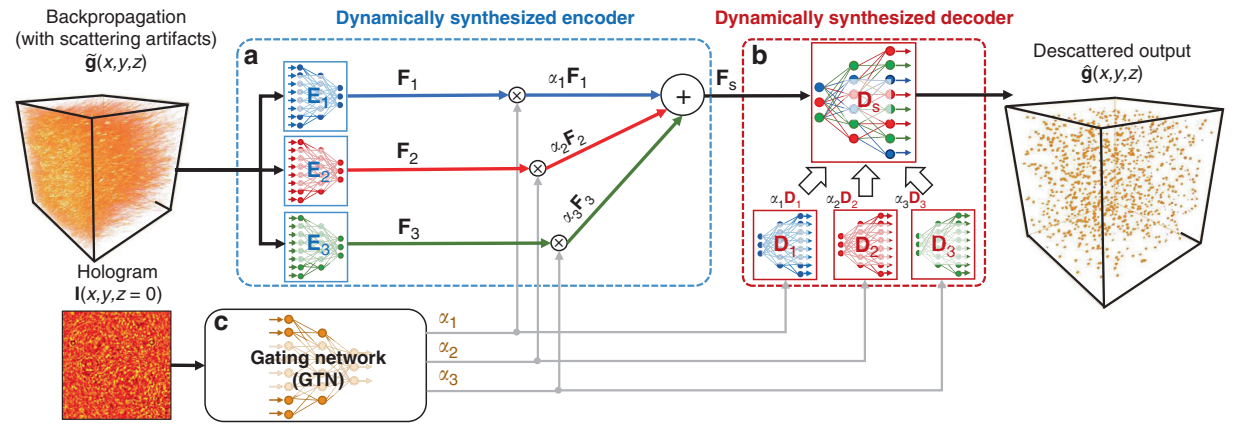

Fig. 1 Dynamic synthesis network (DSN) framework. The DSN combines multiple DNNs for adaptively removing scattering artifacts in the input. **a** In the first stage, the expert encoders E_i ($i \in \{1, 2, 3\}$) extract a diverse set of multi-scale spatial features from the holographically backpropagated input volume $\tilde{g}$. The extracted multi-scale feature maps from each encoder are labeled as F_i. To adaptively process an input with an arbitrary scattering condition, a dynamically synthesized feature map F_s is computed as a weighted sum of the expert feature maps: $F_s = \sum_{i=1}^{3} \boldsymbol{\alpha}_i F_i$. The synthesized feature F_s is fed into a dynamically synthesized decoder D_s to produce the descattered output $\hat{g}$. **b** Different from the encoder, the decoder D_s is computed as a weighted sum of the expert decoders' network parameters: $D_s = \sum_{i=1}^{3} \boldsymbol{\alpha}_i D_i$. **c** The GTN provides the adapting mechanism by predicting the synthesis weights $\boldsymbol{\alpha}_i$ based on the matching hologram input

than the expert since the generalist needs to perform the challenging task of extracting generalizable features across different scattering cases.

Given these limitations, it is highly desirable to develop a unified adaptive DL framework that can robustly handle a broad range of scattering conditions. While this challenge has not been addressed broadly in the literature, one approach[24] is to train a bank of expert DNNs, each for a different scattering condition. This bank is preceded by a separately trained classification DNN, whose purpose is to select a suitable descattering expert based on the input data at the test time. However, this architecture suffers from poor scalability, and its overall performance is fundamentally limited by the disjoint classification and descattering networks, which are separately trained and applied to the input data during the testing.

Another inspiration stems from the framework of "mixture of experts (MoE)", in which a MoE network is fused into a single network to achieve better generalization[28–31]. In image denoising, a MoE framework exploits a weighted sum of the output from different expert denoising DNNs. The weights for combining the experts are obtained either by solving a separate nonlinear optimization problem or training a separate DNN. This approach has shown significant improvement to the denoising DNN's robustness against different noise types (e.g. Poisson vs Gaussian) and noise levels[29,30]. In phase retrieval, a "learning to synthesize" MoE framework is employed to fuse low- and high- spatial frequency information extracted separately from two expert networks and later synthesized by a residual learning approach[32]. This framework enables phase recovery with improved spatial resolution and resilience to high-noise conditions.

In this paper, we present the first, to the best of our knowledge, holistic adaptive descattering DL framework. The proposed DNN, termed dynamic synthesis network (DSN), further advances the MoE framework by a novel expert mixing scheme. At the high level, the DSN consists of multiple expert descattering networks for learning several sets of object/scattering features to achieve optimal descattering. A gating network (GTN) is introduced to adaptively make "central cognitive decisions" for mixing the experts into a unified synthesized network (Fig. 1). Once trained, the DSN adapts to each input during the inference and computes the optimal weights and the resulting synthesized network "on-the-fly", hence achieving "end-to-end" adaptive descattering. Instead of directly mixing the experts' output in the existing MoE models, the DSN architecture adaptively adjusts the network's (internal) parameters. Specifically, the DSN works by continuously mixing the features maps extracted by different experts to synthesize an optimal feature representation of the input in a high-dimensional feature space (Fig. 1). We show that this novel structure allows DSN to perform adaptive descattering in a continuum of scattering levels and achieve superior performance across a broad range of scattering conditions in simulation and experiments.

In this work, we use 3D inline holography as the testing bed to demonstrate our DSN framework. Specifically, we perform 3D descattering of volumetric particle field reconstructions from the single-shot inline holography. We demonstrate the adaptability and robustness of the

DSN over a broad range of scatterer densities, sizes, and refractive index contrast in simulation and further validate the DSN on experimental measurements with different scatterer densities. We show that the DSN can adaptively remove scattering artifacts even if the scattering condition has never been "seen" during the training, and its performance is comparable to, if not even better than, the expert network separately trained at the matching scattering condition. We show that the DSN provides better generalization performance than the generalist network that is trained on multiple-scattering densities simultaneously.

A widely perceived challenge in demonstrating the effectiveness of a supervised DNN model in real optical experiments is the difficulty in capturing a sufficiently large data set with paired input and ground-truth data to train or fine-tune the DNN. This is indeed the case in our intended application, which involves imaging freely moving micron-sized particles in fluid[33]. To overcome this challenge, we apply our recently developed multiple-scattering simulator-trained DNN framework[5]. Specifically, we train the DSN entirely using simulated data, which are generated by the beam-propagation model[34], in place of physically acquiring paired training data. We show that this simulator-trained DSN presents similar behaviors on the experimental data to those observed in simulation and can robustly perform adaptive 3D descattering across a broad range of scatterer densities. By doing so, we verify that the DSN can be successfully translated from simulation to real experiment. We also highlight the synergies between the DSN framework and the physics-based multiple-scattering simulation to enable scientific and biomedical applications where the ground truth is hard to obtain.

Overall, our contribution is a novel adaptive DL framework that achieves generalizable performance across a broad range of scattering conditions using a single holistic DNN architecture. We demonstrated this framework on 3D particle imaging using inline holography with different particle densities, particle sizes, and refractive index contrast. We expect that the same dynamic synthesis framework can be adapted to many other imaging applications, such as image denoising[35], imaging in dynamic scattering media[36,37], computational fluorescence microscopy[38,39], and imaging[13,23–25] and light control[40–43] in complex media, such as biological tissues. Broadly, our dynamic synthesis framework opens up a new paradigm for designing highly adaptive DL-based computational imaging techniques.

Results

In this section, we provide an overview of the DSN framework, followed by results on simulation and experiments.

The DSN framework

The unique properties of the DSN include its dynamically synthesized feature representations of the input and the adaptively tuned network parameters, both of which are adjusted "on-the-fly" at each inference time to achieve adaptation. This is in stark contrast to conventional DNNs, which perform direct inference with pretrained (i.e. fixed) network parameters. The DSN also bypasses the limitations in rigid model switching between a fixed set of expert DNNs, which makes the DSN more versatile and scalable. The DSN enables synthesizing a network in a continuous high-dimensional feature space so that it can provide optimal performance in a continuum of scattering conditions. The adaptability of the DSN stems from the interplay between a GTN and a consortium of expert DNNs. Each expert DNN in the DSN extracts certain spatial features to provide a diverse representation of the input. The GTN provides the feedback needed to intelligently and dynamically fuse the extracted features and synthesize the feature representation tailored to the current input, as further illustrated in Fig. S4. It also enables computing the optimal mixture of the network parameters to utilize the synthesized features for descattering at different conditions.

In this proof-of-concept study, we demonstrate the utility of our DSN framework to perform 3D descattering on holographically backpropagated 3D volumes containing high-contrast densely distributed particles (see details in Section "Experimental setup"). Due to the large scale of the problem, we choose three experts to remove scattering artifacts from the backpropagated volumes, in order to strike a balance between the descattering performance and the computational cost. The effects of the number of expert networks are further studied in Fig. S6.

The schematics of our DSN framework are shown in Fig. 1. The input to the network is a preprocessed scattering-contaminated backpropagated volume from the hologram (see details in "Holographic backpropagation" and "Data preprocessing"). The network is trained to remove scattering artifacts, whose severity is highly dependent on the scatterer density, size, refractive index contrast, and further complicated by the depth-varying characteristics throughout the 3D volume[34,44]. Within the DSN, each expert has the same modified V-net structure[45], which is further split into the expert encoder and expert decoder for performing dynamic synthesis (see Fig. 1, and more details about the network structure in Section "DSN network design and implementation", Figs. S1 and S2).

In the first stage, each expert encoder independently extracts a set of multi-scale spatial feature maps from the scattering-contaminated input volume. Since each expert encoder has different "specializations", the combined set of feature maps provide a diverse representation of the

scattering volume that provides the basis for adaptation. To intelligently utilize these multi-scale features for processing an arbitrary scattering condition, a linearly weighted sum of the extracted feature maps is computed using the "synthesis weights" predicted by the GTN at each inference time (see Fig. 1a). By doing so, this dynamically synthesized encoder provides a set of tailored feature maps to represent the input (see an illustration in Fig. S4a). The training of the DSN effectively optimizes the set of bases (feature maps) in order to get a more generalizable representation for different scattering conditions.

On the decoding end, the DSN dynamically synthesizes the decoder during the inference by mixing a set of expert decoders. This process is performed by directly computing a linearly weighted sum of the network parameters of the expert decoders (see Fig. 1b). Finally, the synthesized features from the encoder are processed by the synthesized decoder to produce the descattered output volume (see an illustration in Fig. S4b).

The GTN provides the "feedback signal" for dynamic synthesis by predicting the synthesis weights (Fig. 1c). These weights can be thought of as the "coefficients" under the scattering "bases" learned by the expert DNNs in the DSN for linear synthesis. To perform this task, we design a GTN that extracts multi-scale spatial features from the matching preprocessed hologram (see more details in Section "Data preprocessing") and outputs three scalar numbers $\{\alpha_1, \alpha_2, \alpha_3\}$ that sum to unity: $\sum_{i=1}^{3} \alpha_i = 1$ (see more details in Section "DSN network design and implementation" and Fig. S3). In this way, the GTN performs a holistic analysis of the level of input scattering artifacts. It enables adaptive mixture of the features extracted by the expert encoders and linear synthesis of the expert decoder parameters. Therefore, the GTN can be considered as the primary agent for adaptability in the DSN.

Training and initialization strategy

During the training, all the expert encoders and decoders in the DSN are co-trained with the GTN (see details in Section "DSN network design and implementation"). In this way, the GTN learns to optimally combine the expert encoders and decoders for varying levels of input scattering. During the same time, the experts (E_i and D_i) are observed to gain "specializations" to different scattering conditions, as discussed in detail later. In order to impart broad adaptability to the DSN, the training data contains a diverse set of scattering levels (see details in Section "Simulated training and testing data sets").

We investigate two DSN initialization strategies. In both cases, the GTN is initialized using the Xavier random weight scheme[46] while the parameters in the expert DNNs are initialized differently. In the first scheme, for which results are shown in the main text, each expert network in the DSN is initialized by first pre-training on a data set with a specific level of scattering density (see details in Section "Simulated training and testing data sets"). This is particularly helpful to illustrate the MoE type of working mechanism of the DSN, as we discuss in detail in Section "Analysis of gating network". In addition, instead of fixing the parameters of the pretrained experts for synthesizing the network, we continue the training process by co-training the expert networks with the random-initialized GTN to obtain the final DSN parameters.

In the second initialization scheme, we remove the pre-training requirement and initialize all the experts and the GTN with Xavier random weights[46] (see details in Section "DSN network design and implementation"). Our results show that the DSN still converges to a multi-expert configuration despite different initialization methods (as shown in Fig. S7). We further demonstrate that regardless of the initialization scheme (i.e. with pretrained or randomly initialized expert parameters), the DSN produces similar descattering performance, as shown in Fig. S8. This highlights that the native DSN architecture can both impart specializations to its constituent experts, and enable adaptability by its inherent dynamic reconfiguration capability.

Performance in seen scattering conditions in simulation

First, we present 3D descattering performance of the DSN on "seen" scattering conditions (i.e., the scattering conditions used in the training). For this purpose, we first simulate inline holograms from high-contrast (refractive index contrast $\Delta n = 0.26$) microspheres (diameter $D = 1.0\,\mu\text{m}$) randomly distributed in a 3D volume ($176.64 \times 176.64 \times 500\ \mu\text{m}^3$). We generate data from four different particle densities ($\rho = \{1.6, 3.2, 6.41, 12.82\} \times 10^4$ particles μL^{-1}). The density was chosen to increase by a factor of 2× in each step in order to cover a broad range of scattering levels, ranging from sparse to dense cases. To accurately model the multiple-scattering effects, we generate holograms using the beam-propagation method (BPM), whose high accuracy has been established experimentally by our recent work[34]. Additional details about the simulation are provided in Section "Multiple-scattering simulation". To perform 3D reconstruction from each hologram using the DNNs, we first perform 3D holographic backpropagation as a prepossessing step (see details in Section "Holographic backpropagation"). The backpropagated volume is contaminated by scattering artifacts, whose severity depends on the particle density, size, refractive index contrast, and generally worsens as the depth increases[44]. The backpropagated volumes, ground-truth volumes, and the corresponding simulated holograms are used for training each DNN (see details in

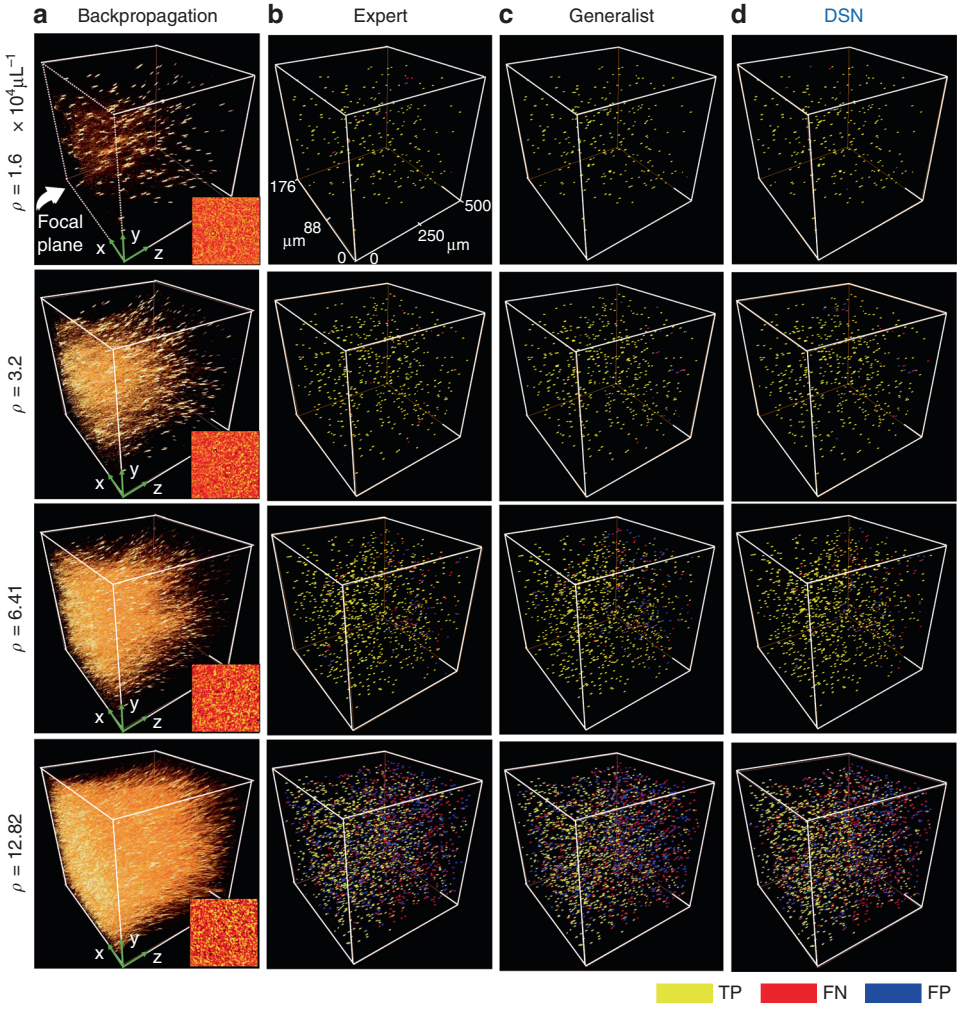

Fig. 2 Particle localization visualization. a 3D renderings of the backpropagated volumes demonstrate the scattering artifacts for various particle densities. The corresponding hologram is shown as the insets. **b**–**d** Particle 3D localization results are shown for the expert, generalist and DSN, respectively. For low particle densities, all three DNNs perform similarly. As the particle density increases, the DSN provides improved performance, measured by the true positives (TP, in yellow), false negatives (FN, in red), and false positives (FP, in blue). The numbers of TP, FP, and FN are quantified in Table S1, highlighting that the results from the DSN contain much fewer FPs at high particle density cases

Sections "Data preprocessing" and "Simulated training and testing data sets").

Next, we apply each trained DNN to remove the 3D scattering artifacts in order to localize the particles. For each density, we first visualize examples of 3D particle localization results of the DSN in Fig. 2 and compare them with two alternative strategies, including the expert network and the generalist network at each matching scattering condition. For this initial comparison, both the expert and the generalist networks use the same V-net architecture and the same number of trainable parameters, as detailed in Section "DSN network design and implementation", Fig. S1 and Table S2. In Section "Comparison of the DSN and generalist", we further benchmark the performance of a "larger" generalist network having approximately the same number of trainable parameters as the DSN and 3× numbers of parameters as the expert network. The generalist network is trained using the same data set as the DSN. To perform DNN testing, preprocessed holographically back-propagated 3D volumes are used as the input, which are computed from holograms never used during the training. The input contains scattering artifacts that increase significantly with the particle density, as visually evident from Fig. 2a. A patch from the corresponding hologram is also shown as the inset in Fig. 2a to demonstrate how the fringe pattern qualitatively varies with the particle density and begins to resemble speckle patterns at high particle densities. In Fig. 2b–d, we visualize the 3D particle localization results by labeling each particle in the network's

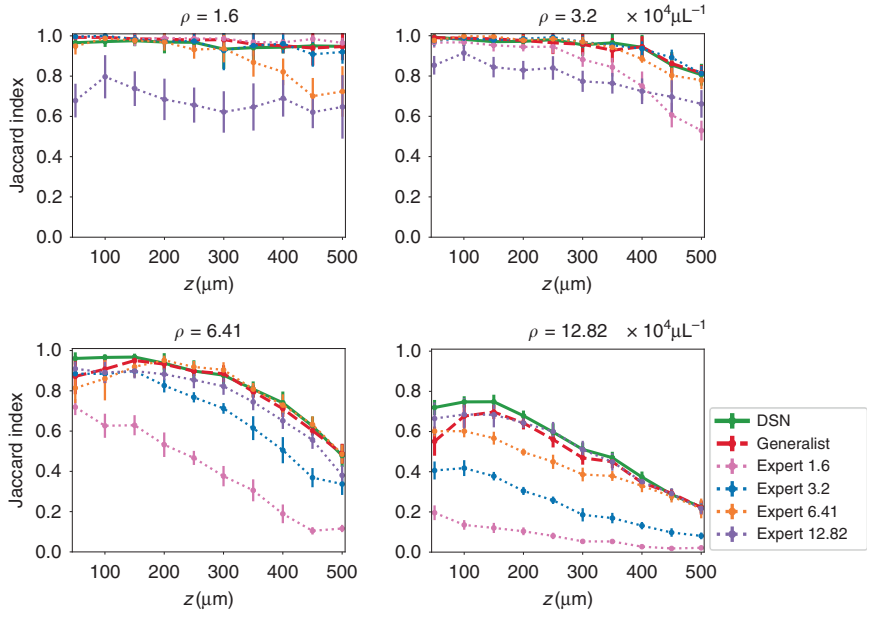

Fig. 3 Quantitative performance evaluation on seen densities. Particle localization performance is quantitatively compared between the DSN, the generalist, and the expert DNNs using JI. Each plot indicates the results on a test data set at the density labeled above each plot and with particle diameter 1.0 μm and refractive index contrast 0.26. "Expert ρ" represents the expert DNN trained on the data set with the density ρ ($\times 10^4$ particles μL^{-1}). The DSN generally provides higher accuracy than both the generalist and the matching expert, in particular for high densities ($\rho \geq 6.41 \times 10^4$ particles μL^{-1})

output by true positive (TP), false positive (FP), or false negative (FN). The labeling procedure is detailed in Section "Performance evaluation metrics". The numbers of TP, FP, and FN particles for each density in the test set (refractive index contrast = 0.26, particle size = 1 μm, 10 volumes) are quantified in Table S1. For relatively low particle densities ($\rho \leq 3.2 \times 10^4$ particles μL^{-1}), all three DNNs demonstrate similar and highly accurate localization. For higher particle densities ($\rho \geq 6.41 \times 10^4$ particles μL^{-1}), the DSN outperforms both the generalist and the matching expert.

A particular challenge of holographic 3D localization is that the scattering artifacts become more severe at greater depth due to the accumulation of coherent diffraction effects from shallower depths[44]. In addition, the initially estimated particles from the holographic 3D back-propagation are elongated more severely at greater depth due to decreasing effective numerical aperture (NA), which aggravates the "missing-cone" artifacts[34], as clearly seen in Fig. S10. These confounding factors imply that even for a fixed particle density, there exists a significant amount of variations in terms of signal fidelity (from the particle) and scattering noise in the input volume. As a result, we expect depth-dependent reconstruction performance. We quantify the performance of the DSN by comparing the estimated and the ground-truth particle locations using the Jaccard Index (JI) similarity score across the depths. We evaluate the depth-wise localization performance by quantifying the statistics of the JI on every 10 axial slices (i.e. every 50 μm) in the reconstructed 3D volumes. Additional details about the quantitative evaluation are provided in Section "Performance evaluation metrics". The same evaluation procedure is also applied to the reconstructions from the generalist and all the expert DNNs. We use "Expert ρ" to denote the expert DNN trained on the data with a specific particle density of $\rho \times 10^4$ particles μL^{-1}. The localization accuracy curves are shown in Fig. 3, in which each plot quantifies the performance of all the DNNs on the same testing data set with the testing particle density labeled above each plot. Each testing data set at a given particle density comprises of ten volumes of particles with the same diameter (1.0 μm) and refractive index contrast (0.26) and different randomly generated 3D locations. Each error bar quantifies the mean and the standard deviation of the JI computed on all the reconstructed particles within the corresponding ten slices of each testing volume and across all the testing volumes.

We generally observe that the DSN outperforms the generalist and the matching expert DNN at high particle densities ($\rho \geq 6.41 \times 10^4$ particles μL^{-1}). Each expert DNN generally performs well on the testing data with the matching density; however, its performance degrades when the testing density is different from the training. Interestingly, both Expert 3.2 and Expert 6.41 provide robust performance for the first three lower density cases

($\rho \leq 6.41 \times 10^4$ particles μL^{-1}). Expert 12.82 provides robust performance for the last two higher density cases ($\rho \geq 6.41 \times 10^4$ particles μL^{-1}). These observations suggest that there indeed exist *generalizable multi-scale spatial features across different scattering densities*. This heuristically provides the foundation to form and combine a set of scattering "bases" for dynamic synthesis in the DSN. Combining all the scattering cases in a brute-force manner for training the generalist network can provide good performance across all the scattering levels, and achieve improved generalizability to different particle densities as compared to the expert DNNs. Such behavior has been observed in our previous works as well[23,25]. However, the generalist generally provides lower accuracy than the expert DNN trained at the matching condition and the DSN.

Additionally, we make the following observations for individual testing densities. For $\rho = 1.6 \times 10^4$ particles μL^{-1}case, the DSN (green solid line) has good performance (JI > 0.9) across all depths, but is slightly worse than the expert (pink dotted line). Upon visual inspection, we attribute this to the following observation. Due to the low particle density and the need to crop the full measurement to small patches, both the training and testing sets contain a small number of "empty" sub-volumes with no particles, i.e., the ground-truth volume is all zero. The DSN tends to "hallucinate" with FPs in these empty regions. The matching expert suffers less from this artifact possibly because its training data set contains approximately 4× more empty volumes than the DSN (see details in Section "Simulated training and testing data sets"). For $\rho = 3.2 \times 10^4$ particles μL^{-1} case, the DSN performs similarly to both the generalist (red dashed line) and Expert 3.2 (blue dotted line). Due to the increased density, all training and testing patches are not empty and hence do not suffer from the issues found in the lower density case. The performance begins to degrade as the particle density increases to $\rho = 6.41 \times 10^4$ particles μL^{-1}, especially at large depths ($z \geq 400\ \mu m$). The DSN outperforms both the generalist (red dashed line) and Expert 6.41 (orange dotted line). The accuracy of Expert 1.6 decreases sharply in this case, which shows that it cannot generalize and handle the increased scattering artifacts. Expert 6.41 suffers from relatively lower accuracy at the first 200 μm shallower depths. We attribute this to the significant variations in the signal-to-noise level throughout the 500 μm imaging volume that arise due to the depth-dependent scattering artifacts at this relatively high density. As seen from the example shown in Fig. S10, at deeper regions, the particles are harder to be distinguished from the background scattering "noise" in the backpropagated volume. Since the expert DNN is trained to handle these variations using a loss function that weighs the prediction error *equally* at all depths (see Section "DSN network design and implementation"), the training process tends to find a balance between the high-noise deep regions and low-noise shallow regions, which results in worse performance at the shallow depths. The same problem is also observed in the generalist when applied to $\rho = 6.41 \times 10^4$ particles μL^{-1} and $\rho = 12.82 \times 10^4$ particles μL^{-1} cases. This issue is much alleviated by the DSN, likely because the DSN dynamically fuses the learned features during the inference, which can provide a better feature representation of the scattering input than the expert and generalist networks. To illustrate that different spatial features are extracted by the expert encoders in the DSN, we show example feature maps for two-particle densities in Fig. S5.

DSN generalizes to unseen scattering conditions

Next, we demonstrate the DSN's ability to generalize to "unseen" scattering conditions on testing data whose particle densities, sizes, and refractive index contrast were never used during the training. To provide a holistic view of the DSN's descattering performance, we show the localization JI as a function of imaging depths for "unseen" conditions (dashed lines) in Fig. 4, along with the JI curves for the "seen" conditions (solid lines, identical to the corresponding curves in Fig. 3). The DSN is trained using only the data from the seen cases with a fixed refractive index contrast 0.26, a fixed particle size 1.0 μm, and the four different particle densities, as detailed in Section "Simulated training and testing data sets".

The generalization power of the DSN is quantified in Fig. 4. Our results demonstrate that the DSN adapts to a wide range of scattering conditions and provides robust descattering performance. First, we test the DSN's generalization by changing either the refractive index, the particle size, or the density. In Fig. 4a, we show that the DSN is highly robust to the change of the refractive index contrast (Δn), yielding negligible localization performance variations for the entire tested Δn range (from 0.01 to 0.50). A possible explanation is that our data normalization procedure (see Section "Data preprocessing") facilitates the DSN to focus on scattering-dependent spatial features by removing the "superficial" changes in the low-order (including mean and standard deviation) intensity statistics (see Fig. S11), and thus provides superior generalization performance to the refractive index contrast. In Fig. 4b, we show that the DSN is also robust to the change of the particle size, especially when the testing size is smaller than the trained size. The performance degrades to JI < 0.8 only when the particle size is increased by 80%. This is expected because the increased particle size results in novel spatial features that are not learned by the DSN that is trained using a fixed particle size. In Fig. 4c, we quantify the DSN's performance at various unseen particle densities. The DSN

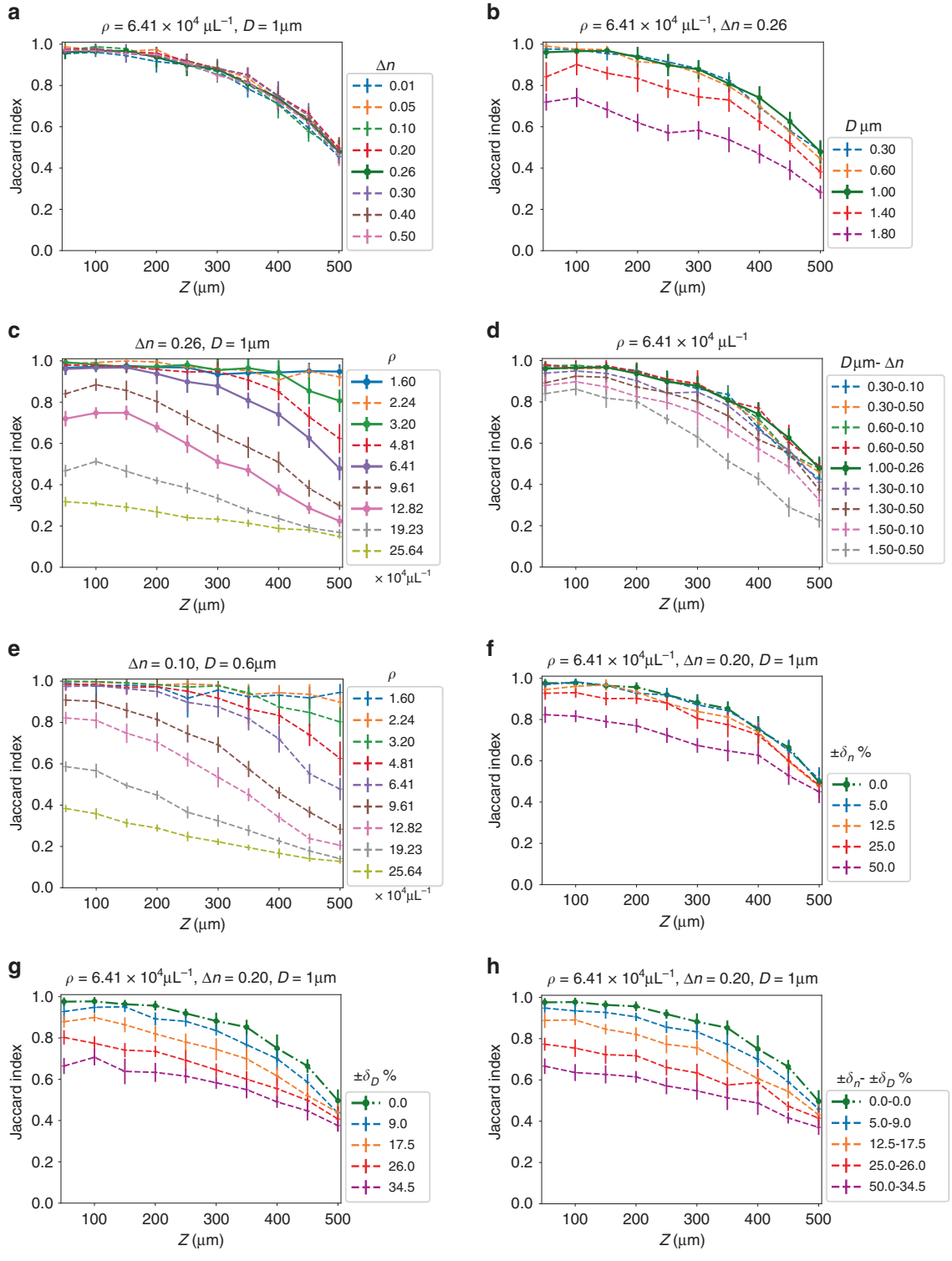

Fig. 4 Generalization of the DSN to unseen scattering conditions. The DSN demonstrates robust descattering for a continuum of scattering levels covering a wide range of refractive indices, particle sizes, and particle densities. The baseline seen cases are shown in solid lines; the unseen test conditions are in dashed lines. The testing cases include: **a** unseen refractive index contrast (Δn); **b** unseen particle size (D); **c** unseen particle density (ρ); **d** unseen refractive index contrast and particle size ($D-\Delta n$); **e** unseen refractive index contrast, particle size and density; **f** Uniformly distributed random refractive index contrast, $\pm\delta_n$% denotes the variation range with respect to the central refractive index contrast; **g** Uniformly distributed random particle size, $\pm\delta_D$% denotes the variation range with respect to the central size; **h** Uniformly distributed random refractive index and particle size ($\pm\delta_n$ to $\pm\delta_D$). In **f**–**h**, the green dash-dotted line is the baseline unseen case at $\rho = 6.41 \times 10^4$ particles μL^{-1}, with a fixed refractive index contrast $\Delta n = 0.20$ and a fixed particle size $D = 1.0$ μm

demonstrates a smooth transition between different scattering levels, regardless of whether they are seen or unseen. This implies that the overall decrease in performance at higher particle densities is more associated with the increased level of scattering artifacts in the input volume (as expected), but less affected by whether the scattering level has been used during the training (i.e., network generalization).

Next, we show the generalization of the DSN on data sets with both unseen particle size and unseen refractive index at a given density in Fig. 4d. Similar to the observations in Fig. 4a, b, the DSN is still robust to these changes. We further test the DSN under a fixed unseen particle size and a fixed unseen refractive index for a variety of particle densities in Fig. 4e. By comparing the results in Fig. 4c, e (which have matching particle densities), we confirm that the DSN's descattering performance is indeed primarily controlled by the particle density.

Finally, we test the DSN on volumes containing particles with randomly distributed refractive indices and sizes. To quantify the effect, we assume both the refractive index and the particle size follow uniform distributions (see details in Section "Simulated training and testing data sets"). The performance of DSN gradually degrades as the variation range increases. In Fig. 4f, the results show that the DSN is robust to random variations of the refractive index up to around 50% amount of fluctuations, beyond which the performance degrades to JI < 0.8 in most of the depths. The fact that the DSN is robust to the global change in the refractive index contrast (Fig. 4a) but degrades on randomly distributed values (Fig. 4f) suggests that the network's performance is limited by the dynamic range in the measurement. Low-contrast particles result in weaker in-focus signals in the backpropagated volume and thus are harder to reconstruct in particular in the presence of high-contrast particles, both of which are surrounded by the same strong background scattering noise. In Fig. 4g, the random variations of the particle size result in reduced accuracy, similar to those observed in Fig. 4b. The results indicate that the DSN is robust to random variations of the particle size up to around 26% amount of fluctuations, beyond which the performance degrades to JI < 0.8 in most of the depths. In Fig. 4h, random variations in both the refractive index and the particle size also lead to smoothly reduced performance. By comparing Fig. 4g, h (without and with particle size variations), we conclude that the generalization of the DSN is primarily affected by the particle size variations.

Overall, these results highlight that the DSN can perform adaptive 3D descattering for a continuum of scattering levels beyond the three specific conditions represented by the experts in its native architecture and the four conditions used in the training data set. This bypasses the need for training a large bank of expert DNNs to handle data with a wide range of scattering conditions. The performance of the DSN can potentially be improved, particularly at high particle densities, by increasing the number of expert networks within the DSN, based on our study in Fig. S6. The performance can also be improved by diversifying the training data to include additional particle densities, sizes, and refractive indices to refine the dynamic synthesis with improved generalization in the future.

Comparison of the DSN and generalist

We further analyze and compare the DSN with two generalist networks (trained on the same data set as the DSN). As described in Section "The DSN Framework" and shown in Fig. S4a, the DSN dynamically combines features extracted by different expert networks to synthesize an optimal feature representation of the input. This means that the synthesized DSN during inference uses the *same* number of feature maps as the expert, as illustrated in Fig. S4b. The DSN achieves this by optimally utilizing the bases learned by the expert networks through a linear feature synthesis. As a comparison, an alternative "brute-force" method to increase the expressing power of a neural network is to directly increase the number of parameters (and hence features), by increasing the number of channels in the convolutional layers and/or increasing the number of layers in the network. Since the number of trainable parameters in an untrained DSN is approximately 3× the number in the expert, we also compare the DSN with a "larger" generalist network containing the same number of trainable parameters, termed "3× Generalist". The hyper-parameters in the 3× Generalist is heuristically optimized. It contains the same number of layers as the expert while having a larger number of convolutional channels in each layer, as detailed in Section "Expert and generalist network implementation" and Table S3.

First, we compare the DSN with the baseline generalist having the same number of feature maps and show that the DSN optimized feature maps indeed result in improved descattering performance. In "Section Performance in Seen Scattering Conditions in Simulation", we have shown that the DSN outperforms the baseline generalist on seen cases (Fig. 3). Next, we compare their generalization capability on unseen scattering conditions. The localization performance of the baseline generalist on the cases in Fig. 4 (shown for the DSN) is provided in Fig. S14. A direct comparison at different unseen densities is also plotted on the same figures in Fig. S12. It is clearly evident that the DSN performs better than the baseline generalist under all different unseen cases.

Next, we compare the DSN with the 3× generalist. First, we compare their performance on seen scattering

conditions in Fig. S13. The DSN and the 3× generalist have similar performance for the first three lower densities ($\rho = \{1.6,\ 3.2,\ 6.41\} \times 10^4$ particles μL^{-1}). The 3× generalist slightly outperforms the DSN at the highest seen density case ($\rho = 12.82 \times 10^4$ particles μL^{-1}), which we attribute to its higher expressing power. Next, we compare the two networks' generalization capability on unseen scattering conditions. Some representative direct comparisons on the most challenging cases containing randomly distributed particle sizes and refractive indices are shown in Fig. 5. Additional results of the 3× generalist for the scattering conditions matching those in Fig. 4 are shown in Fig. S15. The DSN generally performs better than the 3× generalist when the testing condition deviates from the baseline case. These results imply that by increasing the number of feature maps in the 3× generalist, it allows the network to learn more information that is particularly beneficial for descattering the seen high-density cases. However, the larger network also tends to overfit the training data, which results in degraded performance in unseen cases. While it is possible to alleviate overfitting by using a larger training data set, we use the same training set as the DSN for fair comparison. Since the DSN still maintains the same feature map number as the baseline generalist network, it is less prone to overfitting using the same training data set. In addition, the 3× generalist requires higher computational cost than the DSN during inference (more network parameters), as detailed in Sections "The DSN Framework", "DSN network design and implementation", and "Expert and generalist network implementation", which makes the 3× generalist approach less appealing.

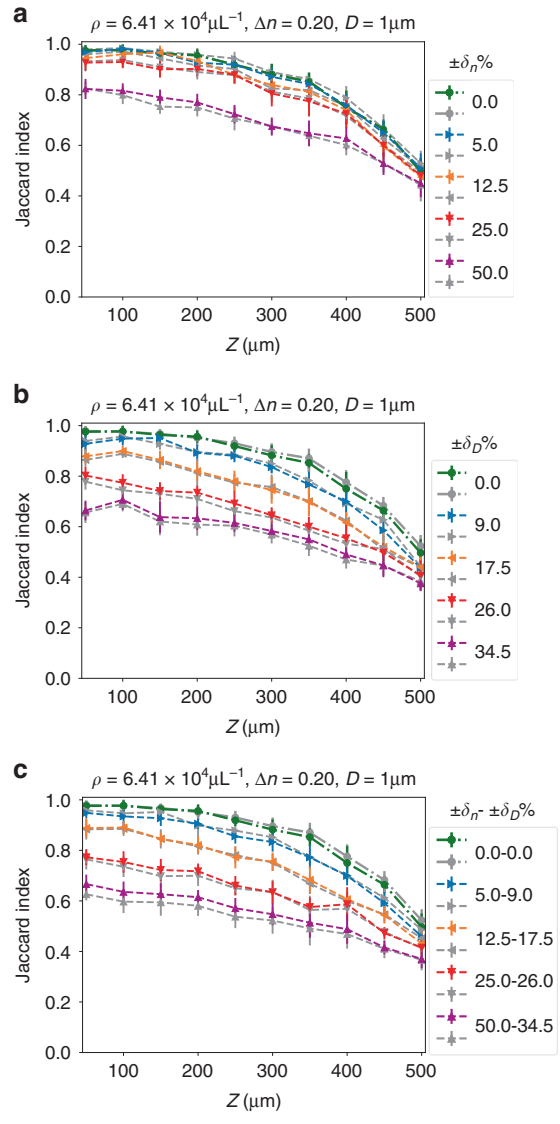

Fig. 5 Comparison of the generalization capability between the DSN and the 3× generalist. The testing volumes contain particles with **a** uniformly distributed random refractive index contrast; **b** uniformly distributed random particle size; and **c** uniformly distributed random refractive index contrast and particle size. The testing particle density (ρ), central refractive index (Δn), and central particle size (D) are shown in the title in each plot. The variation ranges of the refractive indices and the particle sizes are marked by $\pm\delta_n$% and $\pm\delta_D$%, respectively. The green dash-dotted line is the baseline unseen case at $\rho = 6.41 \times 10^4$ particles μL^{-1}, with a fixed refractive index contrast $\Delta n = 0.20$ and a fixed particle size $D = 1.0$ μm. The DSN results are shown in dashed lines in different colors for different cases; the corresponding 3× generalist results are shown in gray dashed lines with the matching markers. The DSN outperforms the 3× generalist as the scattering condition become more different from the baseline case

Analysis of gating network

To elucidate on the working mechanism of the DSN, next we discuss how the GTN dynamically synthesizes the feature maps and network parameters, and in what manner specializations are imparted to the experts within the DSN.

The GTN enables the adaptability of the DSN by generating the synthesis weights for combining the feature maps learned by the expert encoders and parameters of the expert decoders based on the "feedback signal" from the input hologram. It is built on the idea that the raw hologram contains sufficient information to infer the underlying scattering condition. To test this, we use the state-of-the-art unsupervised dimensionality reduction technique, UMAP[47] to visualize the joint distribution of 5120 hologram patches in the training set distributed equally across the four particle densities in Fig. 6a (see more details in Section "UMAP visualization"). For this, we follow the computational procedure established in our recent work[25]. As shown in Fig. 6a, the preprocessed hologram patches with similar underlying particle densities cluster together in the low-dimensional 2D Uniform Manifold Approximation and Projection (UMAP) space. This indicates that the holograms contain

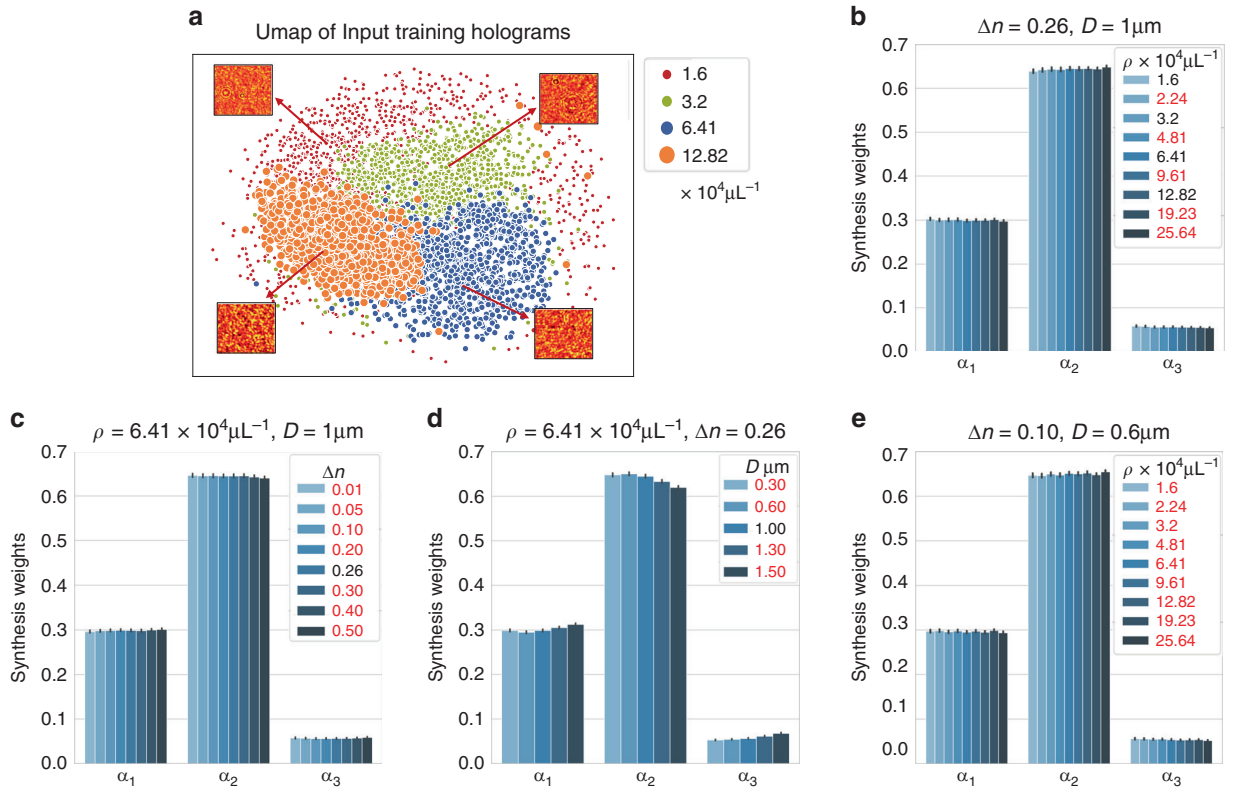

Fig. 6 Gating network analysis. a The 2D UMAP representation for the measured hologram patches used in the training. Hologram patches with similar particle densities cluster together, indicating their statistical similarity. Thus, the hologram can serve as a proxy to the scattering level and is used as the input to the GTN to predict the synthesis weight. The synthesis weights are shown for various seen (legend in black) and unseen (legend in red) scattering conditions, including **b** different particle densities, **c** different refractive indices, **d** different particle diameters, and **e** different particle densities with an unseen particle refractive index and size. The synthesis weights are consistent for each condition and further tailored to each input, as quantified by the mean and standard deviation for each case. The larger values of α_2 indicate the major contributions of F_2 and D_2 to the DSN. As the particle density increases, α_2 increases while α_1 and α_3 decrease regardless of whether the particle size and refractive index are seen (**b**) or unseen (**e**)

intrinsic spatial features to inform the underlying scattering level. This information is extracted by the GTN to adaptively set the synthesis weights for descattering the input volume.

After co-training with the experts within the DSN, the task of the GTN is to *reliably* predict the synthesis weights. A necessary condition is that the GTN should generate similar synthesis weights for holograms measured from similar scattering conditions. To test this, we gather the statistics of the predicted synthesis weights and compute the mean and the standard deviation of each weight α_i for several different scattering conditions, including a wide range of seen (legend in black) and unseen (legend in red) particle densities in Fig. 6b, refractive indices in Fig. 6c, particle sizes in Fig. 6d, and different particle densities at an unseen refractive index ($\Delta n = 0.10$) and an unseen particle size ($D = 0.6$ μm). The results show that the standard deviation is much smaller than the mean for all the cases studied, indicating the consistency of the GTN-predicted synthesis weight. The small weight variations also indicate that the GTN does *not* perform a "rigid" classification of the scattering condition, but rather predicts a set of synthesis weight in a continuous space. It automatically chooses a different linear MoE (i.e. feature maps and decoder parameters) when the scattering condition changes. Thus, by engaging a unique combination of experts for each input, the DSN can not only adapt to different scattering levels but also provide a fine-tuned synthesized network to each input.

Next, we illustrate the different specializations of the experts in a trained DSN. Since the synthesis weights determine the activation of each expert in the synthesized network, we argue that the specialization of each expert can be qualitatively inferred by the statistics of the synthesis weights. A similar interpretation was recently provided by Yang et al.[31]. for analyzing a dynamically synthesized network for robotic locomotion. To demonstrate the expert specializations, we plot the synthesis weights, $\alpha_{i\in\{1,2,3\}}$, for a variety of particle densities (Fig. 6b, e), refractive index contrasts (Fig. 6c), and particle sizes

(Fig. 6d). We note that the overall contribution of α_2 is always the largest compared to α_1 and α_3, making $\mathbf{F}_2$ and D_2 the major contributors of any reconstruction regardless of the input. As the particle density is varied from low to high, α_2 increases while α_1 and α_3 decrease. This trend is consistent regardless of whether the particle size and refractive index are seen (Fig. 6b) or unseen (Fig. 6e). This indicates that the GTN adaptively utilizes more features from $\mathbf{F}_2$ to handle the increased scattering artifacts from the increased particle density. As the refractive index contrast increases for a fixed particle density, the synthesis weights change only slightly (Fig. 6c). Although the increased refractive index results in a greater amount of scattering from individual particles, the data normalization scheme (see Section "Data preprocessing") removes any resulting variations in the low-order statistics (including mean and standard deviation). As a result, the GTN is forced to pay more attention to spatial features induced by inter-particle scattering, which are expected to be less prominent as compared to changing the particle densities. On the other hand, as the particle size increases for a fixed particle density, the changes in the spatial features result in decrease in α_2 and increase in α_1 and α_3. For all the tested data, α_3 remains a small value, indicating that the corresponding expert feature maps $\mathbf{F}_3$ (and the expert decoder D_3) only provide fine-tuned contributions. To ensure that $\mathbf{F}_3$ and D_3 are not redundant, we trained a DSN with only two pairs of expert encoders and decoders in Fig. S6. The two expert DSN under-performs the three expert DSN, which indicates the importance of $\mathbf{F}_3$ and D_3. Overall, a smooth transition is observed for each synthesis weight α_i with respect to the input particle density, refractive index contrast, and particle size regardless of whether they are seen or unseen. This corroborates with our earlier observation of a smooth transition in the DSN performance as the input varies across different scattering conditions.

For the results presented so far, the DSN was trained using the pre-training initialization scheme (see Section "Training and initialization strategy"). Specifically, $E_1(D_1)$, $E_2(D_2)$, $E_3(D_3)$ are first individually pretrained on a specific particle density from 3.2, 6.41, and 12.82×10^4 particles μL^{-1}, respectively (see more details in Section "Simulated training and testing data sets"). We applied this procedure since it is useful to qualitatively associate the specializations of each expert with the particle density used in the pre-training, e.g. $E_1(D_1)$ is specialized in low particle density. However, this leads to a conjecture that the expert specializations in the trained DSN shown in Fig. 6 is merely a consequence of the initial bias introduced by the pre-training. While this might be the case, and even desirable for certain applications[31], we further demonstrate that even with the random initialization, the DSN converges to a state with similar expert specializations. When we perform a similar synthesis weight analysis for the DSN trained using the random initialization scheme in Fig. S7, we can draw a similar conclusion based on our earlier discussions in this section. We note that although different initialization results in different synthesis weight distributions due to the stochastic training process, the observed trend for each synthesis weight remains the same and the trained DSNs produce very similar descattering performance (Fig. S8). Overall, these results again highlight that the DSN architecture natively imparts specializations to its constituent experts, and enables adaptive descattering by dynamically synthesizing the feature maps and adjusting its network parameters.

Generalization of simulator-trained DSN to experimental data

Next, we assess the capability of the DSN on experimental measurements. We capture inline holograms of 3D samples consisting of polystyrene microspheres freely suspended in water, illuminated by a plane wave (more details in Section "Experimental setup"). We use five particle densities, including $\rho = \{1.6, 3.2, 6.41, 12.82, 25.64\} \times 10^4$ particles μL^{-1}, which increases by a factor of 2× in each step to cover a broad range of scattering levels. Following the same procedure as the simulation, each hologram is first backpropagated, preprocessed, and then input to the DSN to perform 3D descattering.

Our results highlight that the *simulator-trained* DSN can be directly used in experiments and robustly perform particle 3D reconstruction across a wide range of densities. We attribute the DSN's generalization capability to two main factors. First, the BPM model we employ to generate the training data in simulation can accurately model the multiple-scattering process (more details in Section "Multiple-scattering simulation"). The statistical properties of the simulated holograms closely match their experimentally measured counterparts, as quantified in our recent work[34]. This means that the input to the DSN does not suffer from "domain shifts", which bypasses the need for transfer learning[48] or domain adaptation[49], thus fulfilling a requirement for the simulator-trained DSN to generalize to experimental data, as also shown in our recent work[5]. Second, the DSN provides good generalization capability to different scattering conditions, which makes it robust to the sample variations present in real experiments, such as particle densities, particle sizes, beam imperfections, and different sources of noise. To demonstrate this, we first plot the synthesis weights of the DSN for experimental data with five different particle densities in Fig. 7. Remarkably, the GTN reliably adjusts the synthesis weights for the experimental measurements in a manner that closely follows the behavior in the simulation (Fig. 6b). The value of each synthesis weight $\alpha_{i\in\{1,2,3\}}$ in the experiment matches well to the simulation

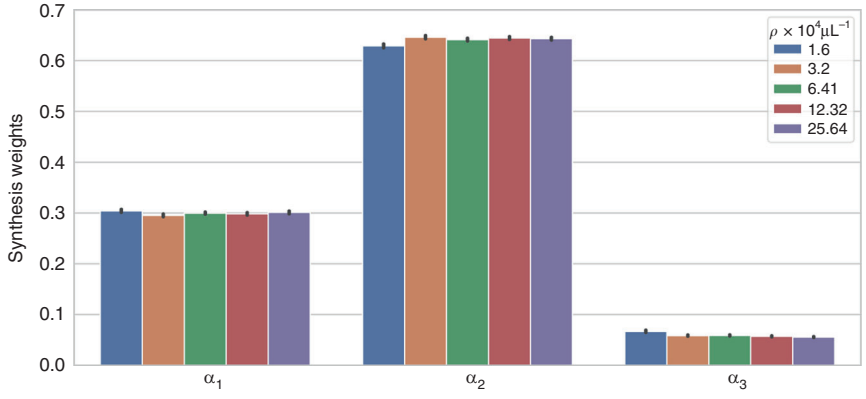

Fig. 7 Synthesis weights of the simulator-trained DSN for experimental data. The synthesis weights computed for the experimentally measured holograms match well with the corresponding simulation results. The mean and standard deviation of each synthesis weight are calculated from 640 non-overlapping hologram patches from 10 experimentally measured holograms for each particle density. This analysis shows that the simulator-trained DSN can robustly adapt to experimental data

at all particle densities. Small variations of the synthesis weights are observed for each particle density, as quantified by the standard deviation on each bar plot in Fig. 7. This shows that the synthesis weights dynamically generated by the GTN are tailored to each scattering condition in the experimental measurements. Similar to the simulation, α_2 has the largest contribution compared to α_1 and α_3 across all the experimentally tested densities. As the particle density increases, α_2 generally increases, while α_1 and α_3 decrease to adapt to the increased particle density.

Next, we compare the localization results from the DSN, the expert, and the generalist networks on the experimentally measured holograms. The results on three representative cases with increasing densities ($\rho = \{1.6, 3.2, 6.41\} \times 10^4$ particles μL^{-1}) are shown in Fig. 8; two additional cases at higher densities ($\rho = \{12.82, 25.64\} \times 10^4$ particles μL^{-1}) are shown in Fig. S16. To perform this comparison, both the expert for a given particle density and the generalist are also trained using only simulated data. For each case, we show the depth color-coded 3D rendering of the particle localization, and the maximum intensity *y*- and *z*-projections overlaid on the corresponding projections of the 3D holographic backpropagation. Although we do not have the ground-truth particle locations for the experiments, the amplitude distribution of the 3D holographic backpropagation can provide visual cues for identifying the particle positions[50], especially at low particle densities. For the lowest density case in Fig. 8a, the expert DNN visibly suffers from misdetections especially at deeper depths, as highlighted by the white ovals in the *y*- and *z*-projections. This is likely because the simulator-trained expert does not have sufficient generalizability to handle the unaccounted-for experimental variations. The DSN and the generalist demonstrate more consistent localization results across the depths. We attribute their superior generalization capability, as compared to the expert, to the increased feature representation power. For the higher density cases in Figs. 8b, c and S16, it becomes challenging to establish the performance for every particle due to the increased scattering artifacts. However, we identify fewer clearly mis-detected regions for the DSN in the *y*- and *z*-projections. Based on these visual inspections, we conclude that our experimental results are consistent with our simulation. The simulator-trained DSN demonstrates robust 3D descattering, and provides improved 3D localization as compared to the expert and the generalist networks, in particular at higher densities and deeper depths.

Discussion

In this study, we have presented and experimentally demonstrated a novel *adaptive* DL framework, termed DSN, for removing volumetric scattering artifacts. We demonstrated the DSN's generalization capability of adaptively removing 3D scattering artifacts and achieving state-of-the-art performance for a wide range of scattering conditions. Broadly, the DSN provides a new "MoE" network architecture, which adjusts the feature maps and network parameters "on-the-fly", to achieve adaptation to different input. Our study also highlighted the utility of multiple-scattering simulator-based training that can enable generalization to real experimental data. This is particularly attractive since the paired ground-truth labels are hard to obtain in many scientific and biomedical imaging applications. In our recent work, we have shown that this simulator-based training strategy can be applied to 3D quantitative phase imaging and achieve high-resolution reconstruction on complex

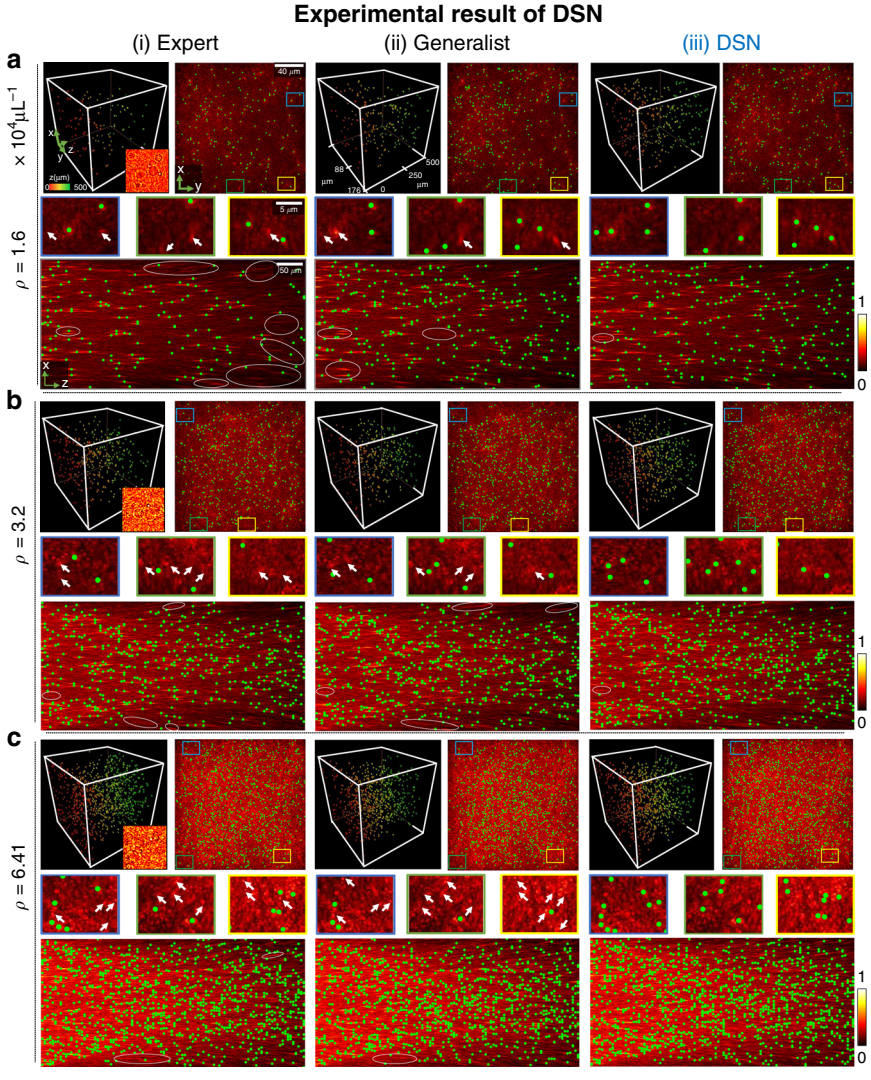

Fig. 8 Experimental results from simulator-trained DNNs. Particle 3D localization is shown for experimentally measured holograms using the simulator-trained (i) expert, (ii) generalist, and (iii) DSN networks at three-particle densities. Each panel shows (Top left) the 3D rendering of the localization result with depth color-coded particles, with an inset showing a zoom-in of the measured hologram, (Top right) the maximum intensity *z*-projection along with three zoom-in regions, and (Bottom) the *y*-projection of the DSN's 3D localization result (in green), overlaid on the respective *y*- and *z*-projections of the corresponding holographic backpropagations. Visually identified mis-detection regions are marked by white ovals in the *y*-projections and white arrows in the *z*-projections

biological samples[5]. We expect the synergy between physics-based large-scale simulation and data-driven models can significantly advance scientific/biomedical DL techniques. Our proof-of-concept demonstration of the DSN focuses on holographic particle 3D imaging, which can find immediate applications in dynamic flow measurements[50–52], imaging cytometry[53–55], and biological sample characterization[56,57]. Broadly, we believe this new adaptive DL framework can be further adapted to many other imaging applications, including image denoising[35], non-line-of-sight imaging[58], deep imaging in scattering tissue[59,60], and computational fluorescence microscopy[38,39].

The DSN structure can be further improved in several aspects in the future. First, using the V-net as the "backbone" to construct the DSN for processing volumetric data poses challenges in scaling up the number of experts in the DSN due to the large computational cost. However, increasing and thus diversifying the expertize is fundamental to push the performance limit of the DSN as shown in this study. Future work may investigate lightweight network structures for processing volumetric data[5,61,62] to reduce the computational cost. Second, in terms of the reconstruction accuracy, the DSN is still limited by its single-scattering approximant backpropagation input, which becomes particularly limiting at

high particle densities. Our recent work has shown that the model-based reconstruction based on the BPM can significantly improve reconstruction quality at strong scattering conditions and large imaging depths[34]. Thus, a promising future direction is to embed multiple-scattering physics into the DSN framework to further push the descattering performance[63]. Third, the DSN's generalization capability can be improved by further diversifying the training set to include different particle size and refractive index variations. A challenge to handle scattering samples with large contrast variations is the limited dynamic range in the measurement that makes the recovery of low-contrast scatterers unstable[5,63]. Novel data augmentation techniques may be explored to overcome this issue[5]. Finally, the feature synthesis process in the DSN is similar to that used in the "attention" mechanism[64]. Thus, it is possible to refine the feature synthesis scheme by using novel attention learning architectures.

Materials and methods

DSN network design and implementation

The expert and baseline generalist networks have the same network architecture (shown in Fig. S1) that is based on the V-net framework[45]. It consists of an encoder and decoder with skip- connections for high-resolution feature forwarding. The encoders and decoders in the DSN (shown in Fig. S2) are also based on the same V-net structure, since they are derived from the expert DNNs. The GTN within the DSN is a relatively simple DNN following the VGG structure[65] to predict the synthesis weights α_i. The detailed implementation of the GTN is shown in Fig. S3. The synthesis weights α_i are used to adaptively mix the extracted feature maps from the three expert encoders and to synthesize the parameters of the synthesized decoder in the DSN, as detailed in Fig. 1.

The DSN framework adaptively combines multiple expert DNNs to achieve dynamic synthesis. In the first stage, three expert encoders $\mathrm{E_i}$ independently perform multi-scale feature extractions from the preprocessed holographically backpropagated volume, where $i \in \{1, 2, 3\}$ represents the expert index. We use $\mathbf{F_i}$ to represent all the extracted multi-scale feature maps from these expert encoders within the DSN, which are then used to compute the dynamically synthesized features $\mathbf{F_s}$ by the following linear combination, $\mathbf{F_s} = \sum_{i=1}^{3} \alpha_i \mathbf{F_i}$. These synthesized features at the corresponding spatial scales are *directly* passed to the synthesized decoder by the skip connections *without* any additional encoding procedures. The synthesized "latent" feature maps (code) at the "bottleneck" (having the lowest dimension $8 \times 8 \times 7$) are passed to the synthesized decoder. Effectively, the latent feature maps represent the adaptively optimized feature representation of the input under the three bases learned by the experts within the DSN. This dynamic synthesis of the encoder is illustrated in Fig. S2.

On the decoding end, the DSN decoder is dynamically synthesized from the multiple expert decoders by $D_\mathrm{s} = \sum_{i=1}^{3} \alpha_i D_i$, where D_s represents the network parameters of the synthesized decoder, and D_i represents the expert decoder parameters within the DSN. Every network parameter in the synthesized decoder is thus a linear combination of the corresponding expert decoder parameters, whose proportions are determined by the GTN. More details about the dynamic synthesis of the decoder are provided in Fig. S2b.

To formulate the training loss function, we treat the ground-truth volume as binary-valued, in which the sample contains discrete particles {1} separated by the background {0}. Accordingly, we employ the binary cross entropy as the loss function, which has shown to promote sparsity in the reconstruction[23]. The loss $\mathcal{L}(\mathbf{W})$ is defined by

$$\mathcal{L}(\mathbf{W}) = -\frac{1}{N}\left(\sum_{i=1}^{N} y_i \log \mathrm{P}(y_i = 1|\mathbf{X};\mathbf{W}) + (1 - y_i)\log \mathrm{P}(y_i = 0|\mathbf{X};\mathbf{W})\right) \tag{1}$$

where y_i is the ground-truth label of a voxel (i.e. 1 for particle, 0 for background), $\mathrm{P}(y_i|\mathbf{X};\mathbf{W})$ is the predicted probability of the voxel belonging to either a particle or the background, given input $\mathbf{X}$ and the network weights $\mathbf{W}$, and N represents the total number of voxels in a given batch.

We use the Adam optimizer to update the network parameters. For training the DSN, while the GTN is initialized with Xavier random weights in all cases, two different initialization strategies are employed for the experts within the DSN, as discussed in Section "Training and initialization strategy". In the first case, each expert within the DSN is initialized with the weight of a pre-trained expert DNN on a specific particle density from $\rho = \{3.2,\ 6.41,\ 12.82\} \times 10^4$ particles $\mu\mathrm{L}^{-1}$, respectively. The pre-training of these expert DNNs follows the same procedure for training the expert networks. With this initialization, the DSN was trained with a learning rate of 10^{-5} with a batch size of 1 for about 30 K iterations. Each iteration takes about 0.62 s, and the training time for 1 epoch is around 5 h. After training, we perform model selection using a validation set. For the expert initialized DSN, the best model is trained for only one epoch. In the second initialization strategy, no pre-training is performed, and all the experts within the DSN are initialized using Xavier random initialization. In this case, the DSN is trained using the same learning parameters as the first initialization strategy. Due to the random initialization, the best model is trained for four epochs, which converges with about 120 K iterations and 20 h. Additionally, L2

weight decay was observed to improve the results from the DSN. Thus, L2 weight regularization was employed for the DSN training, with a regularization parameter $\gamma = 1e^{-6}$, making the effective loss function for the DSN as $\mathcal{L}(\mathbf{W}) + \gamma \mathbf{W}_2^2$.

For our computations, we use an Intel Xeon E5-1630 v4 3.7 GHz processor with 128 GB RAM and an Nvidia Quadro RTX-8000 GPU.

Expert and generalist network implementation

As mentioned in Section "Comparison of the DSN and generalist", we implemented two generalist networks to compare with our DSN. The first generalist network is the same as the expert network, as shown in Fig. S1. The difference is that when training this generalist, the data set contains different scattering data, as detailed in Section "Simulated training and testing data sets". The parameters of each layer in this generalist are detailed in Table S2. We use this generalist as the first baseline because the synthesized DSN combines three different experts into a single V-net as shown in Fig. 1. After the synthesis shown in Fig. S4a, the DSN has the same number of learned features as this generalist (or expert), as shown in Fig. S4b. By comparing this baseline generalist network, we aim to elucidate on how feature synthesis by the DSN improves the generalization capability.

The second generalist network contains 3× number of network parameters, termed 3× generalist, as detailed in Section "Comparison of the DSN and generalist". To implement the 3× generalist, the same V-net structure is maintained while additional numbers of channels are used, as detailed in Table S3. The number of channels used in different layers were heuristically optimized. The 3× generalist provides the second baseline for comparison with our DSN. While they have the same number of trainable parameters in training, during inference, the DSN uses about 3× fewer features to achieve the high performance. This highlights that the dynamic feature synthesis enabled by the DSN structure makes the network more efficient to generalize as compared to increasing the network size in brute-force (i.e., in the 3× generalist).

Both generalist networks use Xavier random weight initialization. For the baseline generalist, the network is trained for about 17 K iterations with a learning rate of 10^{-5} and a batch size of 20. Each iteration takes about 2.90 s and the total training time for this network is 14 h. For the 3× generalist, the network is trained for about 30 K iterations (one epoch) with a learning rate of 10^{-4} and a batch size of 1. The best model is selected based on the validation set. Each iteration takes about 0.43 s and the total training time for this network is around 4 h.

The expert network also uses the Xavier random weight initialization and is trained for about 81 K iterations with a learning rate of 10^{-4} and a batch size of 4. Each iteration takes about 0.80 s, and the total training time for each network is about 18 h.

Both generalist networks and the expert networks are trained using the same loss function in Eq. (1). L2 weight regularization is found to have no effect on their results.

Experimental setup

Our experimental setup for inline holography, shown in Fig. S9, uses a linearly polarized HeNe laser ($\lambda =$ 632.8 nm, 500: 1 polarization ratio, Thorlabs HNL210L) that is collimated for illuminating the sample. A 4 F system with a 20× objective lens (0.4 NA, CFI Plan Achro), and a 200 mm tube lens is used to relay the field onto a CMOS sensor (FLIR GS3-U3-123S6M-C, pixel size 3.45 μm, cover glass removed by Wilco imaging, Inc) for recording the holograms, each containing 1024 × 1024 pixels. The effective lateral pixel size is 172.5 nm, which satisfies the Nyquist sampling requirement. The sample consists of polystyrene microspheres suspended in deionized water, with nominal diameter 0.994 μm ± 0.021 μm and refractive index 1.59 (Thermofisher Scientific 4009 A). The sample is held in a quartz-cuvette with inner dimensions $40 \times 40 \times 0.5$ mm^3. The front focal plane of the objective lens was set slightly outside of the inner wall of the cuvette for hologram recording. A shutter speed of 5 ms was used and found to be sufficiently fast to avoid any motion artifacts from the moving particles. The illumination beam diameter was kept less than the width of the cuvette, while larger than the CMOS sensor area to avoid edge artifacts. With this setup, we acquire holograms for samples with approximate particle densities at $\rho = \{1.6, 3.2, 6.4, 12.82, 25.64\} \times 10^4$ particles μL^{-1}, corresponding to approximately 250, 500, 1000, 2000, 4000 particles in the imaged volume of size $176.64 \times 176.64 \times 500$ μm^3.

Multiple-scattering simulation

Since the ground-truth particle locations for the experimental holograms are not known, we employ simulated data for the DSN training. We simulate 3D samples with particle densities that approximately match the experimental data, including $\rho = \{1.6, 3.2, 6.41, 12.82\} \times 10^4$ particles μL^{-1}, containing 250, 500, 1000, 2000 particles in the simulation volume for our training set and different scattering conditions for our test set, more details about the data set are detailed in Section "Simulated training and testing data sets". The particle locations are placed randomly using the Poisson disk random sampling[66]. The particle diameter D is 1.0 μm and the refractive index contrast Δn between the particle and the background medium (water) is 0.26. The size of

each synthetic 3D sample is $176.64 \times 176.64 \times 500$ μm^3, corresponding to $1024 \times 1024 \times 4222$ voxels. The lateral field of view corresponds to the experimental hologram size. The axial size matches with the internal depth of the cuvette used in the experiment. The lateral step size is chosen to be $\delta x = \delta y = 172.5$ nm, which corresponds to the effective pixel size in the experiment. The axial step size is chosen to be $\delta z = 118.4$ nm, corresponding to $\lambda_m/4$ in order to accurately model multiple scattering[34], where λ_m is the wavelength in the aqueous medium, i.e. $\lambda/1.33$, 1.33 being the refractive index of water. For simulating the holograms given the synthetic particle volume, we use the BPM to accurately model the multiple-scattering process since it demonstrates minimal discrepancy between the measured and simulated holograms and is highly computationally efficient for simulating large-scale data with GPU acceleration[34]. The implementation detail of the BPM along with the accompanying open-source code is provided in ref. [34].

Holographic backpropagation

To generate the input to the DNN, we perform holographic backpropagation on each hologram to obtain an initial estimate of the 3D particle fields. This 3D backpropagation corresponds to the minimum-norm solution for the 2D-to-3D reconstruction problem under the linear first Born approximation model[67]. The 3D holographic backpropagation $\mathbf{R}(x, y;z)$ is computed by numerically propagating the hologram $\mathbf{I}(x, y)$ "backwards" from the hologram plane to the object volume slice-by-slice, as follows:

$$\mathbf{R}(x, y; z) = \mathcal{F}^{-1}\{\mathcal{F}\{\mathbf{I}(x, y)\} \cdot \mathcal{H}(u, v; z)\} \tag{2}$$

where $\mathcal{F}\{\cdot\}$ denotes the 2D Fourier transform and $\mathcal{F}^{-1}\{\cdot\}$ its inverse, x and y the lateral spatial coordinates, z the axial distance, and u and v the transverse spatial frequencies. $\mathcal{H}(u,v;z)$ is the transfer function of the free-space Green's function $G(x, y;z) = \exp(ikr)/r$, where $k = 2\pi/\lambda_m$ is the wave number and $r = \sqrt{x^2 + y^2 + z^2}$, and $\mathcal{H}(u,v;z)$ is computed by taking the slice-wise 2D Fourier transform of the Green's function: $\mathcal{H}(u,v;z) = \mathcal{F}\{G(x, y;z)\}$. For the backpropagation, we set $\delta z = 5$ μm to approximately match the axial resolution of the experimental setup, determined by $\lambda_m/\left(1 - \sqrt{1 - \mathrm{NA}^2}\right) = 5.7\,\mu$m. This is because we do not attempt to localize particles with better accuracy than the system's axial resolution. This produces a backpropagated volume with $1024 \times 1024 \times 100$ voxels. Since our problem is highly ill-posed due to the large dimensional mismatch between the object and measurement domains, the backpropagated volume contains significant scattering artifacts whose severity is directly proportional to the underlying particle density of the sample and increases at deeper depth slices, as depicted in Figs. 2 and S10.

Data preprocessing

All DNN models in this work are trained using supervised learning framework. Therefore, in order to train a network to remove scattering artifacts from the 3D backpropagated volume, the corresponding ground-truth object is also required. The ground truth is difficult to obtain for our experimental data, instead we *only* use simulated data for the DNN training, for which the ground truth is available. A remaining challenge is the large scale of the 3D imaging problem. Each synthetic 3D sample contains $1024 \times 1024 \times 4222$ voxels, which is too large to be computed directly on our DNNs. Simply dividing the depth into smaller sub-volumes would lose contextual information about the depth-dependent scattering artifacts clearly visible in the backpropagated volumes in Fig. 2 and S10. Instead, we down-sample the ground-truth volume with axial binning to form a smaller sized volume containing $1024 \times 1024 \times 100$ voxels. For this purpose, we project each of the 4222 slices in the original synthetic 3D sample to the closest slice within the 100 slices in the ground-truth volume. The resulting $1024 \times 1024 \times 100$ volume is used as the ground-truth label for the DNN training. In addition, we pose the training as a detection problem with a binary ground truth, i.e. every voxel represents whether it belongs to a particle {1} or the background {0}. Essentially, we do not attempt to reconstruct the actual refractive index of each particle but only aim to localize them as a compromise to the severe ill-posedness of the inverse problem.

To preprocess the backpropagated volume before inputting it to the DNN, we first take the amplitude of each complex-valued backpropagated volume. Next, we normalized each volume by subtracting its mean and then dividing its standard deviation: $\tilde{\mathbf{g}} = \left(\mathbf{g} - \mu_g\right)/\sigma_g$, where $\tilde{\mathbf{g}}$ is the preprocessed volume, $\mathbf{g}$ is the original volume, and μ_g and σ_g are the mean and standard deviation of the volume, respectively.

The input to the GTN is a preprocessed hologram. We normalize the hologram by $\tilde{\mathbf{I}} = (\mathbf{I} - \mu_h)/\sigma_h$, where $\tilde{\mathbf{I}}$ is the preprocessed hologram, $\mathbf{I}$ is the original simulated or captured hologram, and μ_h and σ_h are the mean and standard deviation of the hologram, respectively. Two example holograms captured from different scattering densities and the intensity histograms before and after the normalization are shown in Fig. S11. It is evident that, by this hologram normalization scheme, the GTN is forced to extract non-trivial differences in holograms captured at different scattering conditions in order to output distinct synthesis weights.

Simulated training and testing data sets

To train each expert DNN for a specific condition (the refractive index is 0.26, the particle diameter is 1.0 μm, and the particle density is varied for different experts), we randomly generate 48 training and 2 validation pairs of the input backpropagation and the corresponding ground-truth volumes for each expert, each of which is 1024 × 1024 × 100 voxels in size. We further divide each volume to have a smaller lateral size while keeping the axial dimension fixed to make the end-to-end 3D computation feasible. Specifically, the input to the DNN comprises 128 × 128 × 100-voxel sub-volumes cropped from the full backpropagation volumes. The sub-volumes are cropped to have a 64 × 64-voxel lateral overlap among the neighboring patches for training, while no overlap for validation and testing. In total, for each expert DNN, we generated 10800 sub-volumes for training, and 128 for validation.

To train the generalist and the DSN, we generate 70 training and 4 validation pairs of 1024 × 1024 × 100-voxel backpropagation and ground-truth volumes, corresponding to 15750 training, and 256 validation sub-volumes after performing the same volume division procedure. When training the DSN using each sub-volume, the matching preprocessed hologram patch (128 × 128 pixels) is also fed into the GTN for predicting the synthesis weights α. The same data set is used for training the generalist and the DSN for fair comparison of their performance. The particle refractive index contrast is 0.26 and the diameter is 1.0 μm, which is the same as the expert network and the particles used in our experiment. The particle densities ($\rho \times 10^4$ particles μL^{-1}) and the number of volumes (n) at each density used for training the generalist and the DSN include: $\rho = 1.6$ ($n = 10$), $\rho = 3.2$ ($n = 20$), $\rho = 6.41$ ($n = 20$), and $\rho = 12.82$ ($n = 20$). The validation data set includes $\rho = 1.6$ ($n = 1$), $\rho = 3.2$ ($n = 1$), $\rho = 6.41$ ($n = 1$), $\rho = 12.82$ ($n = 1$). We empirically optimized the choice of training data. We chose to not include higher densities like $\rho = 25.64$ for the training, since we observe that the data does not sufficiently benefit the DSN or generalist network training due to the poor effective "signal-to-noise" in the backpropagated volumes.

First, we test all our trained DNNs on a variety of "seen" scattering conditions. We generate 40 testing pairs of backpropagated and ground-truth volumes with 1024 × 1024 × 100 voxels. The refractive index contrast is 0.26, the particle diameter is 1.0 μm, and the four different particle densities are the same as the training set. The particle densities ($\rho \times 10^4$ particles μL^{-1}) used for this testing task and the number of testing pairs (n) include: $\rho = 1.6$ ($n = 10$), $\rho = 3.2$ ($n = 10$), $\rho = 6.41$ ($n = 10$), $\rho = 12.82$ ($n = 10$).

To further test the generalization capability of the DSN and other DNNs, we also test them on several "unseen" scattering conditions, including different particle densities, sizes, and refractive index contrasts. The testing results are summarized in Fig. 4. These testing cases for generalization are categorized into the following eight groups:

(1) Unseen particle densities ($\rho \times 10^4$ particles μL^{-1}) with seen refractive index contrast (0.26) and particle diameter (1.0 μm) and the number of testing pairs (n) include: $\rho = 2.24$ ($n = 10$), $\rho = 4.81$ ($n = 10$), $\rho = 9.61$ ($n = 10$), $\rho = 19.23$ ($n = 10$) and $\rho = 25.64$ ($n = 10$).

(2) Unseen particle diameter (D μm) with seen refractive index contrast (0.26) and density ($\rho = 6.41$) include: $D = 0.30$ ($n = 10$), $D = 0.60$ ($n = 10$), $D = 1.40$ ($n = 10$), $D = 1.80$ ($n = 10$).

(3) Unseen particle refractive index contrast (Δn) with seen particle diameter (1.0 μm) and density ($\rho = 6.41$) include: $\Delta n = 0.01$ ($n = 10$), $\Delta n = 0.05$ ($n = 10$), $\Delta n = 0.10$ ($n = 10$), $\Delta n = 0.20$ ($n = 10$), $\Delta n = 0.30$, $\Delta n = 0.40$ and $\Delta n = 0.50$.

(4) Unseen particle size and Unseen refractive index contrast ($D-\Delta n$) with "seen" particle density ($\rho = 6.41$) include: $D = 0.30-\Delta n = 0.10$ ($n = 10$), $D = 0.30-\Delta n = 0.50$ ($n = 10$), $D = 0.60-\Delta n = 0.10$ ($n = 10$), $D = 0.60-\Delta n = 0.50$ ($n = 10$), $D = 1.30-\Delta n = 0.10$ ($n = 10$), $D = 1.30-\Delta n = 0.50$ ($n = 10$), $D = 1.50-\Delta n = 0.10$ ($n = 10$), $D = 1.50-\Delta n = 0.50$ ($n = 10$).

(5) Unseen particle refractive index contrast ($\Delta n = 0.10$), unseen particle size (0.60 μm) and unseen particle densities ($\rho \times 10^4$ particles μL^{-1}) include: $\rho = 2.24$ ($n = 10$), $\rho = 4.81$ ($n = 10$), $\rho = 9.61$ ($n = 10$), $\rho = 19.23$ ($n = 10$) and $\rho = 25.64$ ($n = 10$).

(6) Volumes containing particles with random refractive index contrasts that follows a uniform distribution $U(\Delta n_c, \delta_n)$ with a fixed central refractive index contrast ($\Delta n_c = 0.20$) and different amounts of variations δ_n, and at a seen particle density ($\rho = 6.41$) and particle size ($D = 1.0$ μm) include: $\delta_n = \pm 5.0\%$ ($n = 10$), $\delta_n = \pm 12.5\%$ ($n = 10$), $\delta_n = \pm 25.0\%$ ($n = 10$) and $\delta_n = \pm 50.0\%$ ($n = 10$), where δ_n is measured by the ratio (%) between the index variation range and the central refractive index contrast.

(7) Volumes containing particles with random diameters that follows a uniform distribution $U(D, \delta_D)$ and different amounts of variations δ_D, and with a fixed central diameter ($D = 1.0$ μm), an unseen refractive index contrast ($\Delta n = 0.20$) and a seen particle density ($\rho = 6.41$) include: $\delta_D = \pm 9.0\%$ ($n = 10$), $\delta_D = \pm 17.5\%$ ($n = 10$), $\delta_D = \pm 26.0\%$ ($n = 10$) and $\delta_D = \pm 34.5\%$ ($n = 10$), where δ_D is measured by the ratio (%) between the diameter variation range and the central diameter.

(8) Volumes containing particles with random refractive index contrasts and random particle diameters, and with a seen density ($\rho = 6.41$). The refractive index contrast follows $U(0.20,\ \delta_n)$. The particle diameter follows $U(1.0\ \mu\text{m},\ \delta_D)$. This testing set, labeled as $\delta_n - \delta_D$, include: $\delta_n = \pm 5.0\% - \delta_D = \pm 9.0\%$ ($n = 10$), $\delta_n = \pm 12.5\% - \delta_D = \pm 17.5\%$ ($n = 10$), $\delta_n = \pm 25.0\% - \delta_D = \pm 26.0\%$ ($n = 10$), $\delta_n = \pm 50.0\% - \delta_D = \pm 34.5\%$ ($n = 10$).

Performance evaluation metrics

The performance of the DNN on the simulated data is quantified using the Jaccard index (JI). To do so, we first discuss the procedure of assigning True Positive (TP), False Positive (FP), and False Negative (FN) labels to the reconstructed particles, and then describe the computation of the JI. TP represents a correctly localized particle, FP represents a falsely detected particle in the reconstruction that is not present in the ground truth, and FN represents a missed particle that is not detected in the reconstruction but is present in the ground truth. The computation involves segmentation of the network output, followed by centroid detection, and the comparison with the ground-truth particle locations, as detailed below.

The output from the DNN is a 3D probability map representing the likelihood of each voxel belonging to either a particle or the background. To process the network's output, first we perform slice-wise automatic thresholding using Otsu's method[68] to obtain a 3D binary map. All clusters composed of less than 10 voxels are discarded as noise, similar to our previous work[44], since they are significantly smaller than the expected particle size of 21 voxels. Second, we perform centroid detection for all the reconstructed particles. Third, we perform pairwise matching between the reconstructed and the ground-truth centroids by computing the pairwise distances and solving a linear assignment problem[69]. Fourth, we assign each reconstructed particle a label from TP, FP, and FN based on the following criterion. TP is assigned to a particle if it is within an elliptical proximity volume of the matching ground-truth particle. We choose the axial dimension of the proximity volume to be 12 μm, which is roughly 2× the axial resolution (5.7 μm) of our system. Even though the lateral resolution of our system is $\lambda_m/\text{NA} = 1.2\ \mu\text{m}$, empirically the localization performance decreases as depth increases. To account for this, we heuristically choose the lateral dimensions of the proximity volume to be 4× the training particle size of 1 μm. Thus, the proximity volume is an ellipse of dimensions $4 \times 4 \times 12\ \mu\text{m}^3$. FP is assigned to a particle if it either does not get matched with any ground-truth particle in the third step, or it is outside the proximity volume of the matching ground-truth particle. FN is assigned to a particle if it is in the ground-truth volume but is not matched with any reconstructed particle in the third step, or the matched reconstructed particle is outside the proximity volume. Finally, using these assigned labels, the JI is computed as

$$\text{JI} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP} + \mathbf{FN}} \tag{3}$$

which measures the similarity between the reconstructed and ground-truth particle locations. The JI is computed for groups of 10 axial slices (i.e., every 50 μm) in the reconstructed 3D volumes.

UMAP visualization

UMAP is the state-of-the-art unsupervised dimensionality reduction technique that models the entire data set into a low-dimensional manifold by learning the underlying topological structure contained in the original high-dimensional data[47]. For UMAP visualization, we consider each data (e.g., a hologram) as a single vector, which is mapped to a single point in a 2D manifold learned by the UMAP algorithm. Applying UMAP to a whole data set can provide insights into the statistical distribution of the high-dimensional data set by visualizing the learned low-dimensional manifold. In the UMAP manifold, statistical similar samples are clustered in close proximity, while dissimilar samples are separated. Its utility for visualizing underlying correlations in scattering measurements have been demonstrated in our recent work[25].

For the UMAP visualization in Fig. 6(a), we use 5120 non-overlapping preprocessed hologram patches of 128 × 128 pixels from 80 preprocessed holograms, distributed equally among the particle densities of $\rho = \{1.6, 3.2, 6.41, 12.82\} \times 10^4$ particles μL^{-1}. From this figure, we can see that even after normalization, the holograms from different particle densities are still clustered based on the underlying density.

Statistical analysis of synthesis weights

We plot the bar chart of output weights given by GTN for different scattering conditions. In Fig. 6b, for particles with refractive index contrast 0.26 and diameter 1.0 μm, we used in-total 1728 non-overlapping hologram patches (128 × 128 pixels) cropped from 27 preprocessed holograms. For each particle density, three preprocessed holograms (corresponding to 192 patches) are used. The following particle densities are used $\rho = \{1.6, 3.34, 3.2, 4.81, 6.41, 9.61, 12.82, 19.23, 25.64\} \times 10^4$ particles μL^{-1}. In Fig. 6c, for particles with density $\rho = 6.41 \times 10^4$ particles μL^{-1} and diameter 1.0 μm, we used in-total 1536 non-overlapping hologram patches (128 × 128 pixels) cropped from 24 preprocessed holograms. For each refractive index contrast, three preprocessed holograms are used. The following refractive index contrasts are

used $\Delta n = 0.01$, 0.05, 0.10, 0.20, 0.26, 0.30, 0.40, 0.50. In Fig. 6d, for particles with density $\rho = 6.41 \times 10^4$ particles μL^{-1} and refractive index contrast 0.26, we used in-total 960 non-overlapping hologram patches (128 × 128 pixels) cropped from 15 preprocessed holograms. For each particle size, three preprocessed holograms are used. The following particle sizes are used $\Delta D = 0.30$, 0.60, 1.00, 1.30, 1.50. In Fig. 6e, for particles with refractive index contrast 0.20 and particle diameter 0.6 μm, we use in-total 1728 non-overlapping hologram patches (128 × 128 pixels) cropped from 27 preprocessed holograms. For each density, three preprocessed holograms (corresponding to 192 patches) are used. The following densities are used from $\rho = \{1.6, 3.34, 3.2, 4.81, 6.41, 9.61, 12.82, 19.23, 25.64\} \times 10^4$ particles μL^{-1}.

Acknowledgements
The authors acknowledge funding from National Science Foundation (1813848 and 1846784). The authors acknowledge Boston University Shared Computing Cluster for proving the computational resources.

Author contributions
W.T. and L.T. conceived the idea. W.T. prototyped the method and conducted initial experiments. H.W. refined the method and conducted additional experiments. All authors participated in the writing of the paper.

Data availability
The neural network and the data set used in this work are available at https://github.com/bu-cisl/DynamicSyntesisNetwork.

Conflict of interest
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41377-022-00730-x.

Received: 30 June 2021 Revised: 22 January 2022 Accepted: 2 February 2022
Published online: 24 February 2022

References

1. Barbastathis, G., Ozcan, A. & Situ, G. H. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
2. Sinha, A. et al. Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).
3. Xue, Y. J. et al. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica* **6**, 618–629 (2019).
4. Wang, F. et al. Phase imaging with an untrained neural network. *Light. Sci. Appl.* **9**, 77 (2020).
5. Matlock, A. & Tian, L. Physical model simulator-trained neural network for computational 3D phase imaging of multiple-scattering samples. Preprint at https://arxiv.org/abs/2103.15795 (2021).
6. Wang, G., Ye, J. C. & De Man, B. Deep learning for tomographic image reconstruction. *Nat. Mach. Intell.* **2**, 737–748 (2020).
7. Liu, J. M. et al. RARE: image reconstruction using deep priors learned without groundtruth. *IEEE J. Sel. Top. Signal Process.* **14**, 1088–1099 (2020).
8. Wu, Z. H. et al. SIMBA: scalable inversion in optical tomography using deep denoising priors. *IEEE J. Sel. Top. Signal Process.* **14**, 1163–1175 (2020).
9. Gupta, H. et al. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE Trans. Med. Imaging* **37**, 1440–1453 (2018).
10. Wang, F. et al. Learning from simulation: an end-to-end deep-learning approach for computational ghost imaging. *Opt. Express* **27**, 25560–25572 (2019).
11. Rizvi, S. et al. DeepGhost: real-time computational ghost imaging via deep learning. *Sci. Rep.* **10**, 11400 (2020).
12. Lyu, M. et al. Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
13. Li, F. Q. et al. Compressive ghost imaging through scattering media with deep learning. *Opt. Express* **28**, 17395–17408 (2020).
14. Wagner, N. et al. Deep learning-enhanced light-field imaging with continuous validation. *Nat. Methods* **18**, 557–563 (2021).
15. Wang, Z. Q. et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nat. Methods* **18**, 551–556 (2021).
16. Wang, H. D. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
17. Liu, T. R. et al. Deep learning-based super-resolution in coherent imaging systems. *Sci. Rep.* **9**, 3926 (2019).
18. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
19. Rivenson, Y. et al. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light. Sci. Appl.* **7**, 17141 (2018).
20. Ren, Z. B., Xu, Z. M. & Lam, E. Y. End-to-end deep learning framework for digital holographic reconstruction. *Adv. Photon.* **1**, 016004 (2019).
21. Rivenson, Y., Wu, Y. C. & Ozcan, A. Deep learning in holography and coherent imaging. *Light Sci. Appl.* **8**, 85 (2019).
22. Wu, Y. C. et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* **8**, 25 (2019).
23. Li, Y. Z., Xue, Y. J. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).
24. Sun, Y. W. et al. Image reconstruction through dynamic scattering media based on deep learning. *Opt. Express* **27**, 16032–16046 (2019).
25. Li, Y. Z. et al. Displacement-agnostic coherent imaging through scatter with an interpretable deep neural network. *Opt. Express* **29**, 2244–2257 (2021).
26. Li, S. et al. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).
27. Sun, Y., Xia, Z. H. & Kamilov, U. S. Efficient and accurate inversion of multiple scattering with deep learning. *Opt. Express* **26**, 14678–14688 (2018).
28. Yuksel, S. E., Wilson, J. N. & Gader, P. D. Twenty years of mixture of experts. *IEEE Trans. Neural Netw. Learn. Syst.* **23**, 1177–1193 (2012).
29. Agostinelli, F., Anderson, M. R. & Lee, H. Adaptive multi-column deep neural networks with application to robust image denoising. In *Proc 26th International Conference on Neural Information Processing Systems*, 1493–1501 (ACM, 2013).
30. Choi, J. H., Elgendy, O. A. & Chan, S. H. Optimal combination of image denoisers. *IEEE Trans. Image Process.* **28**, 4016–4031 (2019).
31. Yang, C. Y. et al. Multi-expert learning of adaptive legged locomotion. *Sci. Robot.* **5**, eabb2174 (2020).
32. Deng, M. et al. Learning to synthesize: robust phase retrieval at low photon counts. *Light Sci. Appl.* **9**, 36 (2020).
33. Katz, J. & Sheng, J. Applications of holography in fluid mechanics and particle dynamics. *Annu. Rev. Fluid Mech.* **42**, 531–555 (2010).
34. Wang, H. et al. Large-scale holographic particle 3D imaging with the beam propagation model. *Opt. Express* **29**, 17159–17172 (2021).
35. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
36. Wang, D. et al. Non-invasive super-resolution imaging through dynamic scattering media. *Nat. Commun.* **12**, 3150 (2021).
37. Zheng, S. S. et al. Incoherent imaging through highly nonstatic and optically thick turbid media based on neural network. *Photon. Res.* **9**, B220–B228 (2021).
38. Pégard, N. C. et al. Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517–524 (2016).
39. Xue, Y. J. et al. Single-shot 3D wide-field fluorescence imaging with a computational miniature mesoscope. *Sci. Adv.* **6**, eabb7508 (2020).
40. Turpin, A., Vishniakou, I. & Seelig, J. D. Light scattering control in transmission and reflection with neural networks. *Opt. Express* **26**, 30911–30929 (2018).
41. Rahmani, B. et al. Actor neural networks for the robust control of partially measured nonlinear systems showcased for image propagation through diffuse media. *Nat. Mach. Intell.* **2**, 403–410 (2020).
42. Turpin, A. et al. Spatial images from temporal data. *Optica* **7**, 900–905 (2020).
43. Skarsoulis, K., Kakkava, E. & Psaltis, D. Predicting optical transmission through complex scattering media from reflection patterns with deep neural networks. *Opt. Commun.* **492**, 126968 (2021).

44. Tahir, W., Kamilov, U. S. & Tian, L. Holographic particle localization under multiple scattering. *Adv. Photon.* **1**, 036003 (2019).
45. Milletari, F., Navab, N. & Ahmadi, S. A. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In *Proc 4th International Conference on 3D Vision (3DV)* 565–571 (IEEE, 2016).
46. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proc 13th International Conference on Artificial Intelligence and Statistics*, 249–256 (JMLR, 2010).
47. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at https://arxiv.org/abs/1802.03426 (2018).
48. Tan, C. Q. et al. A survey on deep transfer learning. In *Proc 27th International Conference on Artificial Neural Networks*, 270–279 (Springer, 2018).
49. Ganin, Y. et al. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**, 2096–2030 (2016).
50. Tian, L. et al. Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography. *Appl. Opt.* **49**, 1549–1554 (2010).
51. Hinsch, K. D. Holographic particle image velocimetry. *Meas. Sci. Technol.* **13**, R61–R72 (2002).
52. Chen, N., Wang, C. L. & Heidrich, W. Snapshot space–time holographic 3D particle tracking velocimetry. *Laser Photon. Rev.* **15**, 2100008 (2021).
53. Cheong, F. C. et al. Flow visualization and flow cytometry with holographic video microscopy. *Opt. Express* **17**, 13071–13079 (2009).
54. Seo, S. et al. Lensfree holographic imaging for on-chip cytometry and diagnostics. *Lab Chip* **9**, 777–787 (2009).
55. Merola, F. et al. Tomographic flow cytometry by digital holography. *Light Sci. Appl.* **6**, e16241 (2017).
56. Moon, I. et al. Automated three-dimensional identification and tracking of micro/nanobiological organisms by computational holographic microscopy. *Proc. IEEE* **97**, 990–1010 (2009).
57. Su, T. W., Xue, L. & Ozcan, A. High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories. *Proc. Natl Acad. Sci. USA* **109**, 16018–16022 (2012).
58. Faccio, D., Velten, A. & Wetzstein, G. Non-line-of-sight imaging. *Nat. Rev. Phys.* **2**, 318–327 (2020).
59. Badon, A. et al. Smart optical coherence tomography for ultra-deep imaging through highly scattering media. *Sci. Adv.* **2**, e1600370 (2016).
60. Kang, S. et al. Imaging deep within a scattering medium using collective accumulation of single-scattered waves. *Nat. Photon.* **9**, 253–258 (2015).
61. Huang, L. Z. et al. Recurrent neural network-based volumetric fluorescence microscopy. *Light Sci. Appl.* **10**, 62 (2021).
62. Kang, I., Goy, A. & Barbastathis, G. Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views. *Light Sci. Appl.* **10**, 74 (2021).
63. Goy, A. et al. High-resolution limited-angle phase tomography of dense layered objects using deep neural networks. *Proc. Natl Acad. Sci. USA* **116**, 19848–19856 (2019).
64. Roy, A. G., Navab, N. & Wachinger, C. Recalibrating fully convolutional networks with spatial and channel "squeeze and excitation" blocks. *IEEE Trans. Med. Imaging.* **38**, 540–549 (2019).
65. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at https://arxiv.org/abs/1409.1556 (2014).
66. Bridson, R. Fast Poisson disk sampling in arbitrary dimensions. In *Proc ACM SIGGRAPH 2007 Sketches*, 22-es (ACM, 2007).
67. Chen, W. S. et al. Empirical concentration bounds for compressive holographic bubble imaging based on a Mie scattering model. *Opt. Express* **23**, 4715–4725 (2015).
68. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
69. Burkard, R. E. & Çela, E. *Handbook of Combinatorial Optimization: Supplement Volume A* (eds. Du, D. Z. & Pardalos, P. M.) 75–149 (Springer, 1999).