






# Artificial confocal microscopy for deep label-free imaging

Received: 27 October 2021

Accepted: 1 December 2022

Published online: 12 January 2023

 Check for updates

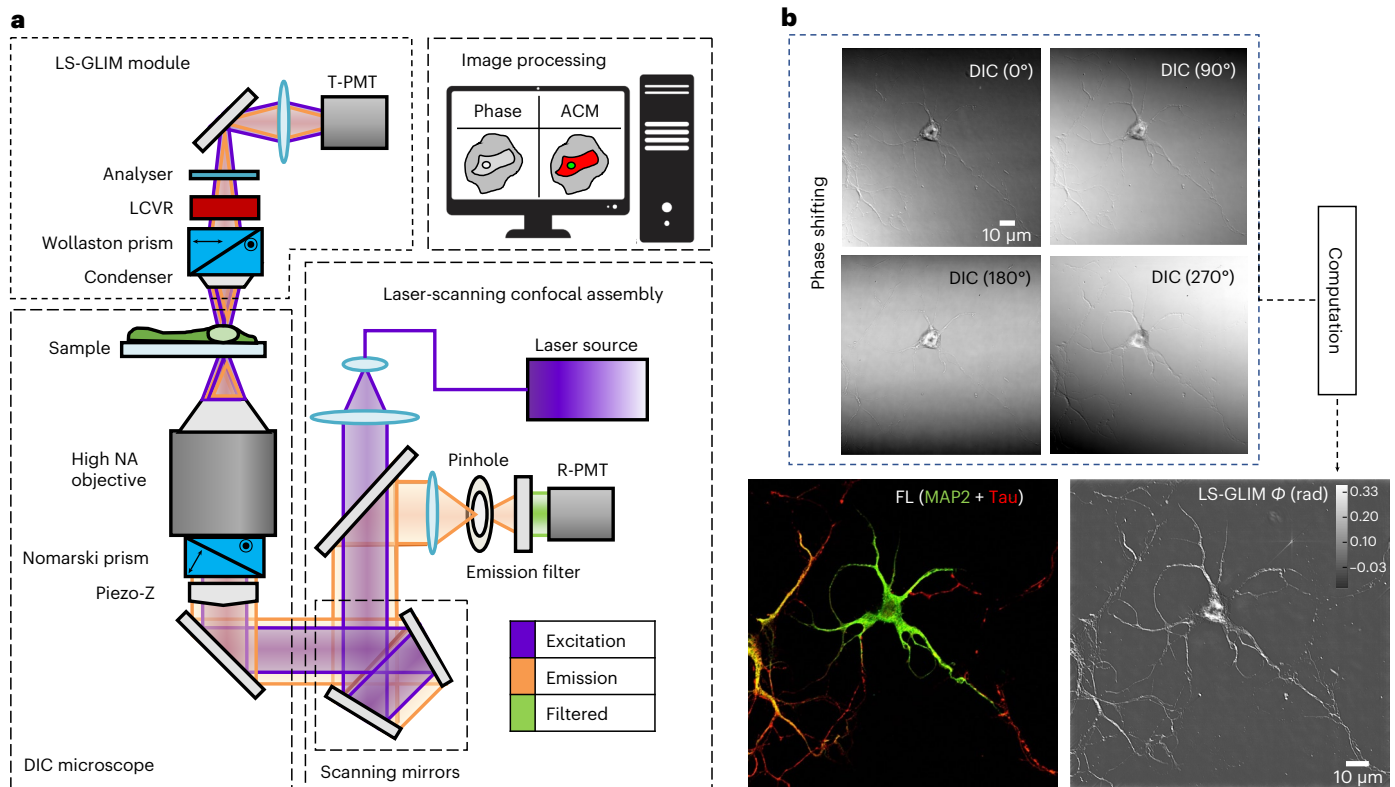
Xi Chen <sup>1,9</sup> ✉, Mikhail E. Kandel<sup>1,10</sup>, Shenghua He <sup>2</sup>, Chenfei Hu <sup>1,3</sup>,  
Young Jae Lee <sup>1,4</sup>, Kathryn Sullivan<sup>5</sup>, Gregory Tracy<sup>6</sup>, Hee Jung Chung<sup>1,4,6,7</sup>,  
Hyun Joon Kong <sup>1,5,7,8</sup>, Mark Anastasio<sup>1,5</sup> & Gabriel Popescu <sup>1,3,5,7</sup>

Wide-field microscopy of optically thick specimens typically features reduced contrast due to spatial cross-talk, in which the signal at each point in the field of view is the result of a superposition from neighbouring points that are simultaneously illuminated. In 1955, Marvin Minsky proposed confocal microscopy as a solution to this problem. Today, laser scanning confocal fluorescence microscopy is broadly used due to its high depth resolution and sensitivity, but comes at the price of photobleaching, chemical and phototoxicity. Here we present artificial confocal microscopy (ACM) to achieve confocal-level depth sectioning, sensitivity and chemical specificity non-destructively on unlabelled specimens. We equipped a commercial laser scanning confocal instrument with a quantitative phase imaging module, which provides optical path-length maps of the specimen in the same field of view as the fluorescence channel. Using pairs of phase and fluorescence images, we trained a convolution neural network to translate the former into the latter. The training to infer a new tag is very practical as the input and ground truth data are intrinsically registered and the data acquisition is automated. The ACM images present much stronger depth sectioning than the input (phase) images, enabling us to recover confocal-like tomographic volumes of microspheres, hippocampal neurons in culture, and three-dimensional liver cancer spheroids. By training on nucleus-specific tags, ACM allows for segmenting individual nuclei within dense spheroids for both cell counting and volume measurements. In summary, ACM can provide quantitative, dynamic data, non-destructively from thick samples while chemical specificity is recovered computationally.

Three-dimensional (3D) cellular systems have been increasingly adopted over 2D cell monolayers to study disease mechanisms<sup>1</sup> and discover drug therapeutics<sup>2</sup>, as they more accurately recapitulate the *in vivo* cellular function and development of extracellular matrices<sup>3</sup>. Three-dimensional cellular structures, including cellular clusters such as organoids and spheroids, have found use in a wide range of applications such as tissue engineering<sup>4</sup>, high-throughput toxicology<sup>5</sup> and personalized medicine<sup>6</sup>. A particularly exciting direction of research

is engineering multicellular living systems<sup>7–9</sup>. These fields of current scientific interest bring along the urgent need for new methods of investigation to inform on cellular viability and cell cluster proliferation. Such techniques would ideally provide quantitative data with subcellular resolution at arbitrary depths in the cellular system and dynamic information rendered over broad time scales. Importantly, these assays would be completely non-destructive, that is, they would report on the cell cluster without interfering with its viability and function.

A full list of affiliations appears at the end of the paper. ✉e-mail: [xc289@cornell.edu](mailto:xc289@cornell.edu)



**Fig. 1 | ACM optical path and image processing.** **a**, The ACM system consists of a laser scanning confocal assembly, a DIC microscope and a LS-GLIM module. QPI was conducted with the green laser line (488 nm) of the confocal excitation. The interferogram was recorded at each point in the scan by the transmission-PMT (T-PMT). The fluorescence images were captured by the reflection-PMT (R-PMT)

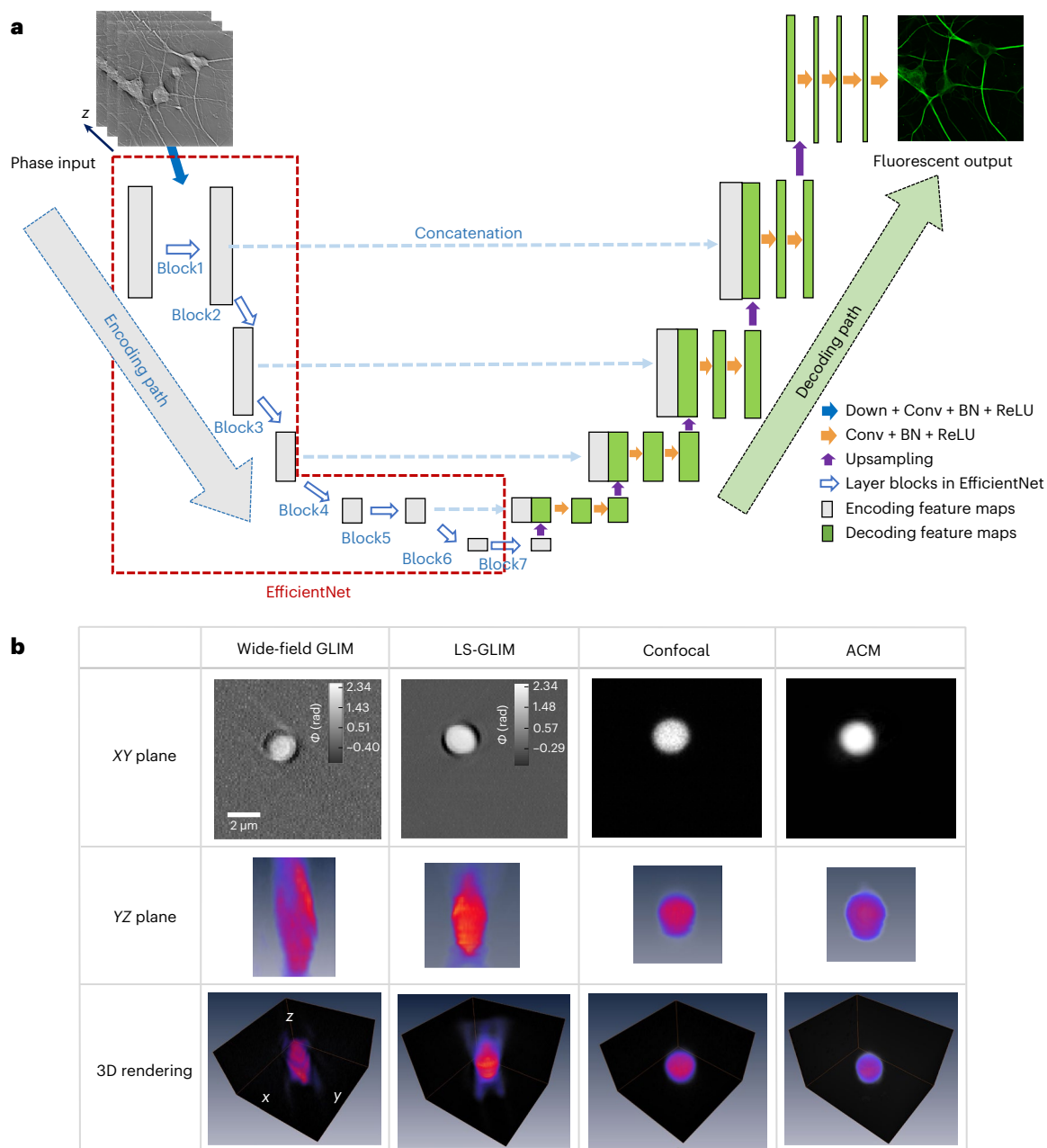
of the confocal module. **b**, Four phase-shifting frames are recorded and used to reconstruct the quantitative phase image. The confocal fluorescence image (FL) serves as the ground truth, whereas the phase image (LS-GLIM) is the input for the network training. NA, numerical aperture.

Due to visible light wavelengths being comparable in size to sub-cellular structures, optical methods of investigation are well suited for meeting these requirements; however, considerable challenges exist for the existing optical microscopy techniques when applied to increasingly thick samples. Typical spheroids—ranging from hundreds of micrometres to millimetres in size—are much larger than the scattering mean free path associated with the light wave propagation, generate strong multiple scattering and therefore form optically turbid aggregates that are difficult to analyse at a cellular level<sup>10,11</sup>. As a result, high-throughput investigations often are limited to extracting coarse parameters, such as spheroid diameters, at low-magnification<sup>12</sup>.

In 1955, in his pursuit to image 3D biological neuronal networks and mimic their behaviour computationally, Minsky was faced with the challenge of suppressing multiple scattering, which was particularly severe for the wide-field instruments available at the time<sup>13</sup>. In Minsky's own words, "One day it occurred to me that the way to avoid all that scattered light was to never allow any unnecessary light to enter in the first place. An ideal microscope would examine each point of the specimen and measure the amount of light scattered or absorbed by that point"<sup>13</sup>. This first implementation of the confocal scanning microscope was established in a transmission geometry, requiring sample translation. Of course, today's modern confocal instruments take advantage of bright laser sources, use beam scanning and are most often used in a reflection geometry, paired with fluorescence contrast<sup>14</sup>. In time, many other advanced laser scanning techniques have been developed for fluorescence microscopy<sup>15</sup>. Nevertheless, fluorescence imaging is subject to several limitations. Absorption of the excitation light may cause the fluorophore to photobleach, which limits the time interval over which continuous imaging can be performed<sup>16</sup>.

The excitation light is typically toxic to cells, a phenomenon referred to as phototoxicity, whereas the exogenous fluorophores themselves can induce chemical toxicity<sup>17</sup>. Although the advancement of green fluorescent protein technology substantially improves the viability of the specimen under investigation, concerns regarding phototoxicity, photobleaching and functional integrity of the cells following genetic engineering still remain<sup>18</sup>. Overcoming these limitations becomes extremely challenging when imaging thick objects over an extended period of time and, for that reason, confocal microscopy is often used on fixed specimens<sup>19,20</sup>.

Multiphoton techniques, including harmonic generation and two-photon (intrinsic) fluorescence microscopy, have been established as valuable label-free approaches for deep-tissue imaging with cellular resolution. Multiphoton microscopy uses excitation light with a longer wavelength that penetrates deeper into tissues, whereas the nonlinear process requires a multiphoton interaction that renders 3D localized excitation<sup>21</sup>. However, multiphoton microscopy requires expensive instrumentations such as femtosecond lasers that are less accessible to the broader community, and the higher-order nonlinear excitation is more susceptible to focus aberrations and phototoxicity. Light sheet fluorescence microscopy can acquire 3D tomography of biological specimens in seconds with high optical sectioning and axial resolution due to the sheet-like illumination, minimizing the background fluorescence and photobleaching<sup>22</sup>. However, tomographic reconstruction often requires sample rotation and sophisticated sample mounting. On the other hand, optical coherence tomography—an interferometric label-free method—was reported to detect and count aqueous cells in the anterior chamber of a rodent model of eye inflammation<sup>23</sup> and volumetrically quantify tumour spheroids<sup>24</sup>. Several



**Fig. 2 | ACM network architecture and inference.** **a**, Network architecture for translating phase images into confocal fluorescence signals. It is a U-Net variant that uses an EfficientNet as the encoder. The input of the Efficient U-Net consists of three adjacent quantitative phase images along the  $z$ -axis, and its output is the corresponding middle fluorescent slice. **b**, Comparison of  $2\ \mu\text{m}$

bead ( $63\times/1.3$ ) tomograms in wide-field GLIM, LS-GLIM, confocal and ACM, as indicated. The elongation of the beads in wide-field and LS-GLIM is due to the missing frequencies in the transmission geometry. On the other hand, the predicted ACM images replicate the confocal sectioning and resolution. BN, batch normalization. Conv, convolution. ReLU, Rectified linear unit.

phase-sensitive methods developed in a confocal modality have been recently developed, but their application to thick structures has been mostly unexplored<sup>25–28</sup>.

Quantitative phase imaging (QPI)<sup>29</sup> has recently emerged as a potentially valuable label-free approach which, due to its high resolution and sensitivity, has found a broad range of new applications<sup>30</sup>. Although most applications involve thin specimens (cell monolayers, thin tissue slices)<sup>31–33</sup>, several efforts have been made for thick multiple-scattering samples. A multilayer Born model using a first Born approximation at each of many layers has been proposed to provide phase reconstruction for thick samples<sup>34</sup>. Nevertheless, the maximum thickness presented in this model was around  $30\ \mu\text{m}$ , and

the performance for 3D organoids with thicknesses usually on the order of  $100\ \mu\text{m}$  remains unclear. A multiscale reconstruction and stitching algorithm for optical diffraction tomography was recently developed to render refractive index distributions of  $100\text{-}\mu\text{m}$ -thick colon tissues with subcellular resolution<sup>35</sup>. The epi-mode tomographic QPI method shows phase reconstructions of cerebral organoids and whole mouse brains with thicknesses of up to  $60\ \mu\text{m}$  via deconvolution<sup>36</sup>. Gradient light interference microscopy (GLIM)<sup>37,38</sup> has been developed to suppress multiple scattering via white light, phase-shifting interferometry, which allowed for imaging and analysis of quantitatively opaque structures such as spheroids and embryos with thicknesses of around  $300\ \mu\text{m}$ . However, as a wide-field technique, GLIM has limited axial

resolution and suffers from spatial cross-talk, which mixes diffraction contributions by neighbouring points from within the specimen. As a result, the accurate discrimination of cellular boundaries deep within a spheroid remains challenging.

Recent developments in artificial intelligence (AI) and machine learning have brought new opportunities to tackle these challenges. Deep learning enables super-resolution in fluorescence microscopy by training a generative adversarial network to transform confocal microscopy images to stimulated emission depletion images, as well as to transform total internal reflection fluorescence microscopy images to the total internal reflection fluorescence microscopy-based structured illumination microscopy images<sup>39</sup>. A deep neural network can be trained to virtually refocus a 2D fluorescence image onto 3D volumetric imaging without any axial scanning, additional hardware, or a trade-off of imaging resolution and speed. This framework is also capable of 3D focusing a single wide-field fluorescence image to match confocal microscopy images at different focal planes<sup>40</sup>. An AI-based deep learning algorithm was developed for the automated quantification of the corneal sub-basal nerve plexus for the diagnosis of diabetic neuropathy using corneal confocal microscopy images<sup>41</sup>. AI-assisted adaptive optics methods aim to compensate systematic and tissue-induced aberrations for imaging deep into turbid specimens<sup>42</sup>. Label-free prediction of 3D confocal fluorescence images can be obtained from either transmitted-light microscopy images or electron micrograph inputs<sup>43</sup>. A label-free, volumetric and automated assessment method has been developed for immunological synapse using optical diffraction tomography and deep learning-based segmentation<sup>44</sup>.

In this Article, we report the artificial confocal microscopy (ACM), a laser scanning QPI system combined with deep learning algorithms, which renders synthetic fluorescence confocal images from unlabelled specimens. First, we developed a laser scanning QPI system, which is implemented as an upgrade module onto an existing laser scanning confocal microscope (LSM 900, Airyscan 2, Zeiss). We validated the boost in sensitivity and axial resolution of the new system by using standard samples and rigorous comparison with the wide-field counterpart. Second, we derived a theoretical model based on the first-order Born approximation, which yields an analytic solution for the spatial frequency coverage of the laser scanning QPI system. These results were validated using experiments to measure the transfer function of the instrument. Third, we trained an artificial neural network on pairs of laser scanning QPI and fluorescence confocal images from the same field of view. As the QPI module is attached to the same optical path, generating the training data is straightforward and automated, as the fields of view are intrinsically registered. Fourth, we applied the inference of the computational neural network to monolayers of biological neural networks and found that the resulting 3D images mimic very well those of the ground truth from the confocal fluorescence images. Using these ACM images, we created binary masks for the contour of the cell and applied them back to the QPI (input) data. Our results show that the measurements of cell volume and dry mass of ACM versus confocal agree very well. Fifth, we used the ACM images to perform nuclear segmentation and, thus, cell counting, within hepatocyte spheroids. We also showed that the training performed on spheroids suspended in phosphate-buffered saline (PBS) can transfer to specimens suspended in hydrogel, which promises broad applications in tissue engineering.

## Results

The ACM imaging system consists of an existing confocal microscope augmented by a laser scanning GLIM system (LS-GLIM). Figure 1a illustrates the ACM set-up, which has three main modules: the LSM (LSM 900, Zeiss), the differential interference contrast (DIC) microscope and the LS-GLIM module. The LS-GLIM assembly shares the laser source from confocal microscopy (see Methods). The two sheared beams that form the DIC image have their relative phase shift controlled by the liquid crystal variable retarder (LCVR), which was carefully calibrated

to produce accurate phase shifts, as described in Supplementary Note 1. For each  $\pi/2$  phase shift, the transmitted light photomultiplier tube (PMT) records the resulting interferogram, as shown in Fig. 1b. The quantitative phase images are generated by the phase-retrieval reconstruction and Hilbert integration algorithms described in the GLIM operation<sup>38</sup>. By sharing the same illumination path, the imaging system registers QPI-z-stack images and pairs them with confocal fluorescence frames from the same field of view, which serve, respectively, as input and ground truth data for the deep learning algorithm (Fig. 1b,c). Due to the laser scanning illumination and PMT detection, the noise level is reduced by a factor of five compared with the full-field method (see Supplementary Note 2); the spatial sensitivity of the phase images is thus improved. The ground truth data (that is, confocal fluorescence images) provide specificity with a high axial resolution and signal-to-noise ratio (SNR). Our goal is to use deep learning to infer the fluorescence confocal images from the LS-GLIM input data and thus replicate the confocal advantages on unlabelled specimens.

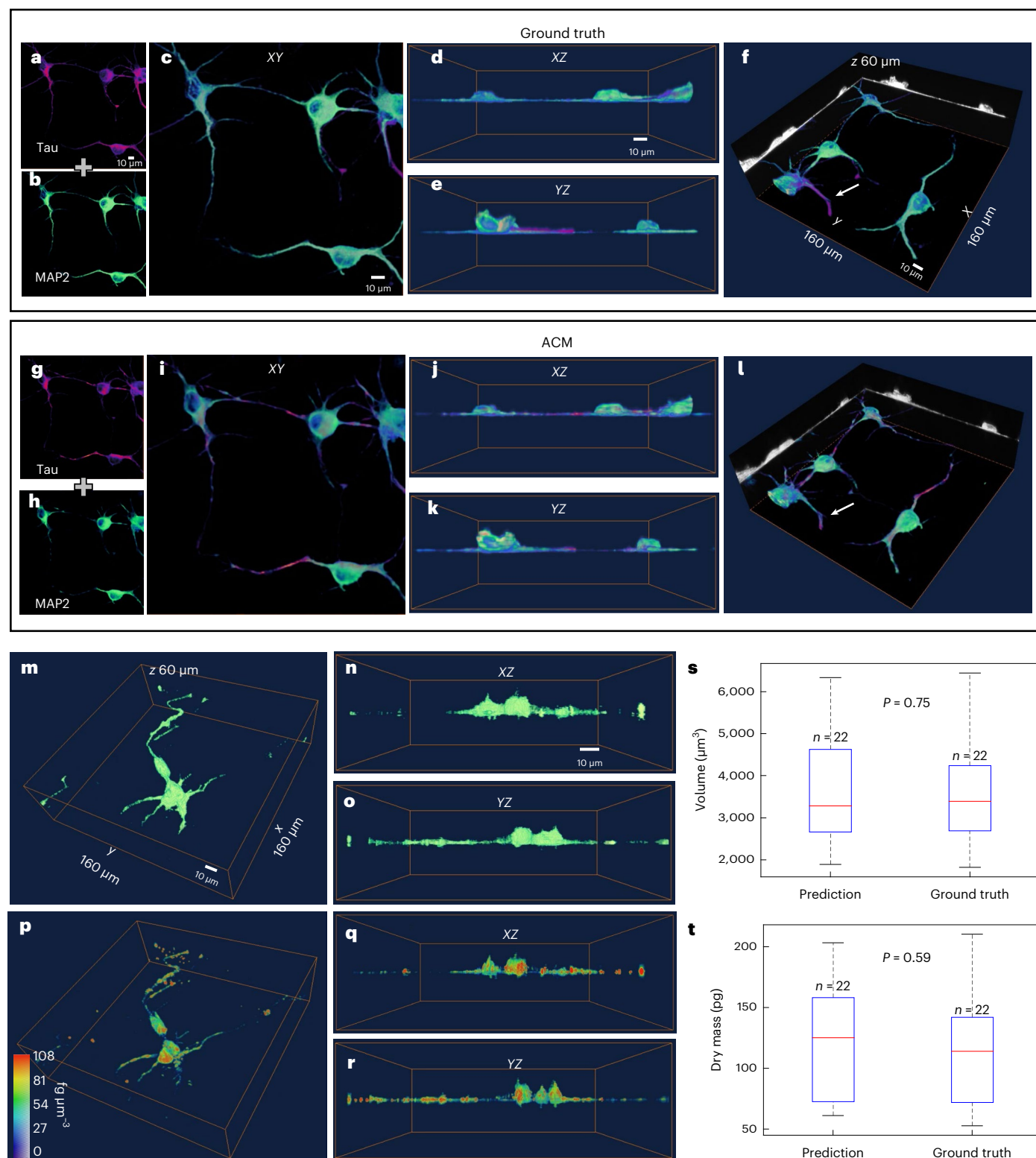
Multichannel EfficientNet-based U-Nets (E-U-Nets) were trained to translate the 3D phase image stack to the corresponding 3D fluorescent image stack. An E-U-Net comprises a standard U-Net where the encoder is replaced with an EfficientNet<sup>45</sup> (Fig. 2a). The multichannel inputs of an E-U-Net are three neighbouring quantitative phase images along the z-axis, and the output is the corresponding central fluorescent image slice (see Methods). We chose this three-frame set as input to account for the fact that the axial spread in LS-GLIM data is much more pronounced than in the confocal fluorescence data, primarily because the input image is obtained in a transmission geometry without a pinhole, whereas the output is in reflection with a pinhole or Airyscan detector array; thus, the neural network learns the spread mechanism from the three adjacent images and reverses it to produce a sharp ACM frame.

In Supplementary Note 3 we present a full description of the 3D image formation in LS-GLIM for weak-scattering samples, which starts with the inhomogeneous wave equation and considers scattering under the Born approximation<sup>46</sup>. The expression for the signal collected at the detector has a particularly simple and physical intuitive form,

$$s(\boldsymbol{\rho}) \propto \chi(\boldsymbol{\rho}) \odot [U_d(\boldsymbol{\rho}) U_i^*(\boldsymbol{\rho})], \quad (1)$$

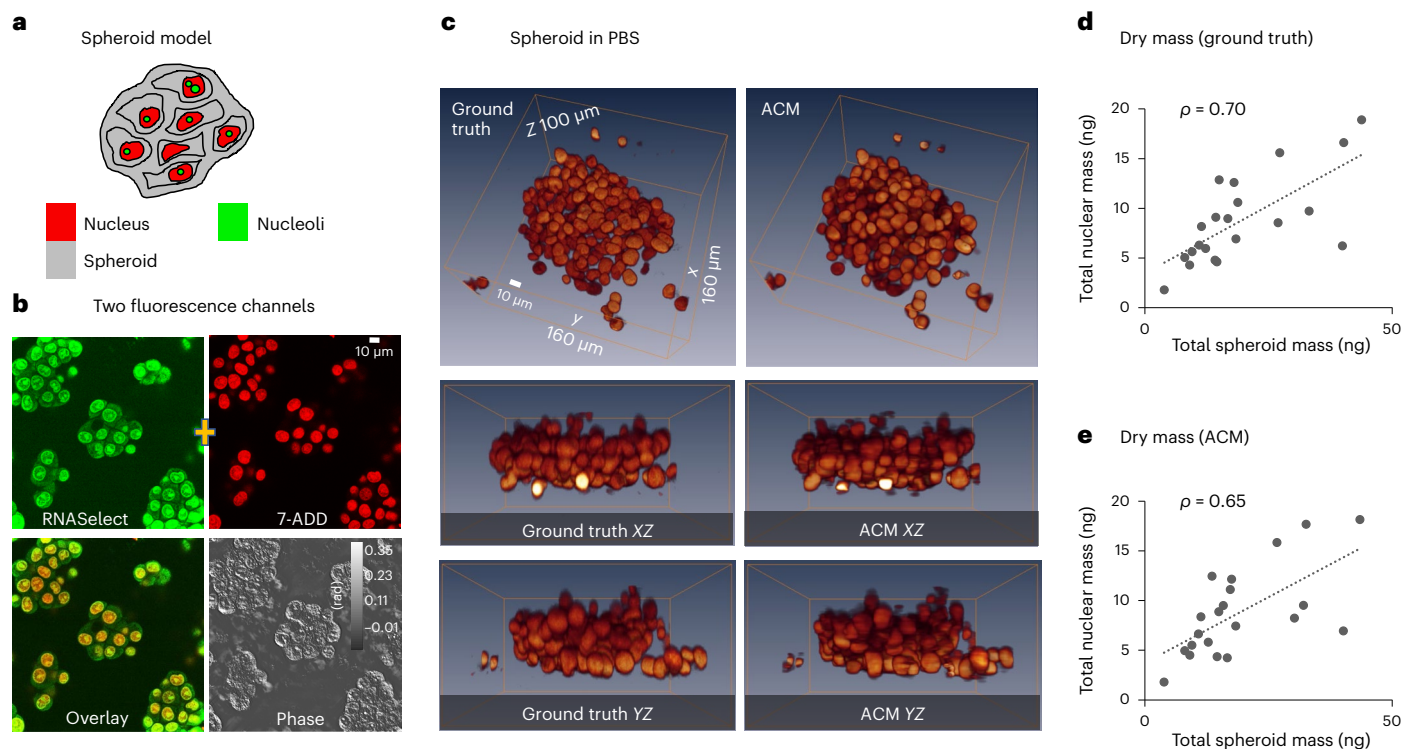
where  $\chi$  is the scattering potential of the specimen;  $U_d$  and  $U_i$  are the detection and illumination functions, defined as the Fourier transforms of their respective pupil functions; the asterisk represents complex conjugation and  $\odot$  denotes the 3D convolution in the spatial domain,  $\boldsymbol{\rho}$ ; the point spread function is therefore given by the product  $U_d(\boldsymbol{\rho}) U_i^*(\boldsymbol{\rho})$ , that is, it improves with both a tighter illumination focus and a broader detection pupil. Note that equation (1) is restricted to weakly scattering specimens such as the phase edge used to estimate the LS-GLIM coherent transfer function (see Supplementary Figs. 5 and 6). These theoretical predictions are comparable with the experimental measurements for various detection numerical apertures. As LS-GLIM uses elastic scattering and operates in transmission mode, the frequencies in the missing cone region of the coherent transfer function cause inferior axial resolution and sectioning compared to confocal fluorescence microscopy. Hence, we rely on the neural network with confocal fluorescence images as reference.

Figure 2b compares images of a 2  $\mu\text{m}$  microbead under wide-field GLIM, LS-GLIM, confocal fluorescence microscopy and the network inference, that is, the ACM image. The resulting ACM image is characterized by its substantially lower axial blur compared with the LS-GLIM input. As described in the Supplementary Note 2, the sensitivity of LS-GLIM is superior to its wide-field counterpart due to the absence of spatial cross-talk and more sensitive photon multiplier detector. However, due to the transmission geometry, they are both inferior to the reflection confocal images in terms of axial sectioning. By contrast, the corresponding network inferences (that is, the ACM images) show much improved axial resolution and sectioning. The average Pearson



**Fig. 3 | ACM estimates volume and dry mass from inferred fluorescence signals. a, b, g, h.** Two-dimensional comparisons of ground truths from confocal (Tau (a), MAP2 (b)) and predicted fluorescence (Tau (g), MAP2 (h)). **c–f, i–l.** A 3D comparison of ground truth from confocal (XY (c), XZ (d), YZ (e), 3D tomogram (f)) and predicted fluorescence (XY (i), XZ (j), YZ (k), 3D tomogram (l)). **m–o.** Volumes from binarized ACM-predicted MAP2 (3D rendering (m), XZ (n), YZ (o)). **p–r.** Dry mass density distribution (3D dry mass density (p), XZ (q), YZ (r)) based

on ACM-predicted MAP2 segmentation and LS-GLIM phase images. **s, t.** The box plots of volume (s) and dry mass (t) for a single cell from confocal MAP2 (ground truth) and ACM predictions (prediction) for DIV neurons. Each box plot shows the median (red lines), the 25th and 75th percentiles of the sample (the bottom and top of each box) and the range (whiskers). *P*-values are of the unpaired two-sided *t*-tests.



**Fig. 4 | Label-free intracellular segmentation in turbid spheroids. a, b**, Cellular compartments were stained using RNA- and DNA-sensitive stains. DNA is used to identify the nucleus and dense concentrations of RNA inside the nucleus are associated with nucleoli. **c**, Three-dimensional comparisons of the ground truth and ACM-predicted tomography of a spheroid (x40/1.3). For all twenty

spheroids, we calculated the nuclear dry mass and volumes generated from the imputed signal. **d, e**, Total nuclear dry mass (ground truth **(d)**, ACM **(e)**) tracks closely with total spheroid mass; PCC  $\rho = +0.65$  (the slope of the linear fit is 0.42). These results agree well with the same procedure applied to the ground truth confocal images **(d)**.

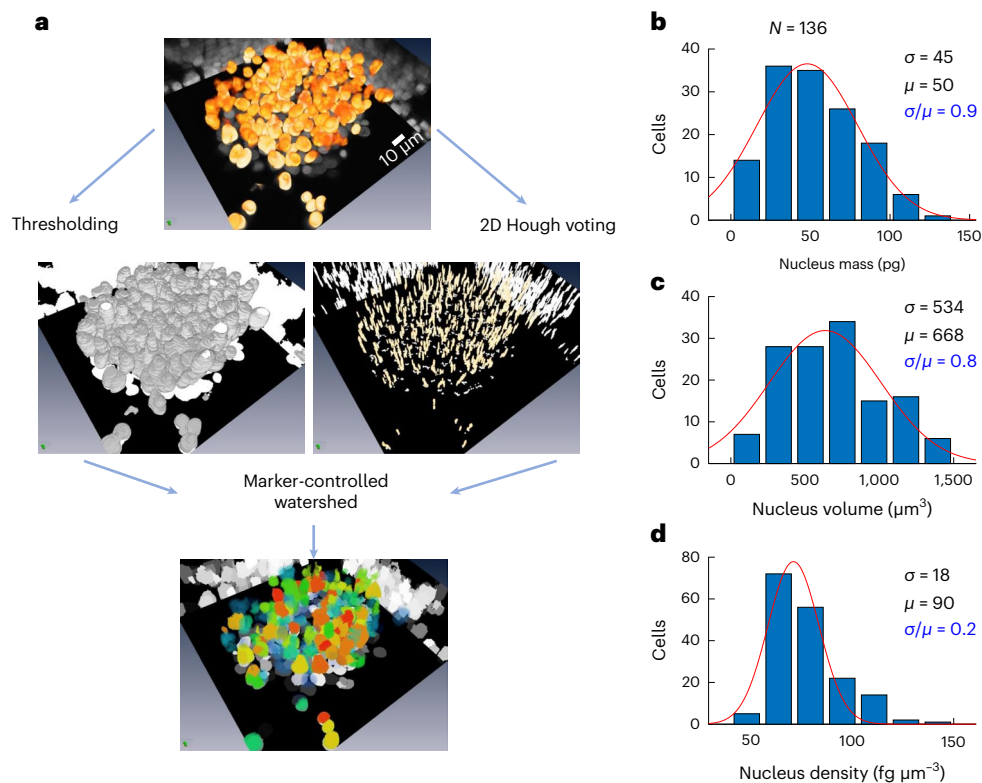
correlation coefficient (PCC) of the microbead is 98% and the peak signal-to-noise ratio (PSNR) is 46.3. The three adjacent LS-GLIM frames used as network input (see Methods for details) contain information about the field Laplacian along  $z$ , which governs the inhomogeneous wave equation (Supplementary Note 3) and may explain why this network architecture can produce adequate results in terms of 3D reconstructions.

We next applied ACM to imaging neural cultures. We used two common stains to tag the Tau and MAP2 proteins<sup>47</sup> (see Methods), the ratios of which are a popular model for differentiating the long axon from smaller dendrites. The confocal fluorescence images from the two channels represent the ground truth and, as before, the corresponding LS-GLIM images were the input data. The training data contained 20  $z$ -stacks of neurons at 10 days in vitro (DIV 10). The results are summarized in Fig. 3a–l. Our results indicate that the overall 3D renderings of the ground truth and their inferences match very well. For Fig. 3a–l, the PCC and PSNR of channel Tau are 80% and 26.9, respectively, whereas they are 91% and 29.1, respectively, for MAP2. We occasionally found some discrepancies in the dendrites, which translates into lower correlation values. The white arrows in Fig. 3f, l point to the axon of the neuron. ACM images reduced the pixel-level noise and confocal stripe artifacts present in the training data. In Extended Data Fig. 1 we compared the power spectra of the neurons from Fig. 3a–l using ground truth and ACM images. The 3D frequency coverage of the ground truth and ACM spectra agree, and both reach the theoretical confocal fluorescence resolution limits. The ACM data allow us to delineate individual cells accurately and measure their volumes. Supplementary Video 1 illustrates this performance on live neurons that have never been labelled. Visually, it is evident that the ACM provides a much sharper decay of the out-of-focus light (that is, greater depth

sectioning) than the original LS-GLIM. Supplementary Video 2 illustrates the time-lapse performance of ACM on unlabelled, dynamic neurons. Of course, the ACM images do not suffer from bleaching or toxicity while maintaining chemical specificity through computation. As a result, ACM is suitable for studying live cellular systems non-destructively over large periods of time.

From the ACM images, we computed binary masks corresponding to the cell contours, which were applied back to the input QPI maps to retrieve individual cell dry mass values. From the cell volume and mass, we also extracted the dry mass density for each cell. The volumes in Fig. 3m–o are rendered using binarized ACM-predicted MAP2, and the dry mass densities in Fig. 3p–r are calculated from ACM-predicted MAP2 segmentation and LS-GLIM phase images. The PCC and PSNR of channel MAP2 are 90% and 32.8, respectively, for Fig. 3m–r. Figure 3s, t shows comparisons of the volume and dry mass associated with a single cell measured from confocal MAP2 and ACM MAP2 predictions. The average volume of a single cell is determined by the total volume per field of view divided by the number of cell body within the field of view. Our results indicate that the volume and dry mass measurements are well-matched with the ground truth, that is, there is not a significant difference between the two distributions ( $P$ -value  $\gg 0.05$ ).

To demonstrate ACM's ability to delineate cellular structures inside turbid spheroids, we imaged hepatocyte spheroids (HepG2) suspended in PBS and generated computational stains associated with the DNA and RNA (Fig. 4a). The RNA is localized within the nucleus, with a high concentration in the nucleolus (Fig. 4b). The study of RNA is currently of high interest, not only as it plays a crucial role in catalysing cellular processes, but also as it can be used by various viruses to encode their genetic information<sup>48</sup>. The two ground truth stains (7-aminoactinomycin D (7-ADD) and SYTO RNASelect



**Fig. 5 | Automated segmentation of cells inside spheroids.** **a**, Instance segmentation of spheroids was performed by 3D marker-controlled watershed on the ACM-estimated DNA signal, with markers being determined through 2D Hough voting on a per-*z*-slice basis. The result of the Hough voting is a volume with a unique marker on the spheroid, which resembles a column tracking the centre of the nucleus through the focus. The result of watershed is a 3D volume

with a unique label for each nucleus within the spheroid, which enables the calculation of parameters on individual cells. **b**, Distribution of nuclear dry mass. **c**, Distribution of nuclear volume. **d**, Distribution of dry mass density;  $\sigma$ ,  $\mu$  and their ratios are indicated for each plot. Note that the nuclear mass density (**d**) has a much narrower distribution than the nuclear mass (**b**) and volume (**c**), as indicated by the  $\sigma/\mu$  ratio.

Green; see Methods) and their associated inferences enable us to generate semantic segmentations and annotate the spheroid into nuclei and nucleoli, respectively. The entire spheroid represents our third class and is obtained as the non-background regions in the LS-GLIM data. As shown in Fig. 4c and Supplementary Note 4, the actual and imputed fluorescent maps show good agreement. The PPC and PSNR of channel DNA are 84% and 24.9, respectively. As detailed in Methods, we apply a threshold on the ACM and phase image to generate 3D semantic segmentation maps, which we use to measure the dimension of the spheroid. The intersection of the RNA and DNA labels provides the annotation for the nucleoli. Our results show that the total nuclear mass is proportional to the spheroid mass across the twenty spheroids studied in this work. This dependence is shown in Fig. 4d, where the slope of the linear regression (0.42) indicates that about 42% of the spheroid mass is contributed by the nuclei. These results agree well with the same procedure applied to the ground truth confocal images.

Automatic instance segmentation of cells inside spheroids was performed by 3D marker-controlled watershed on the estimated DNA signal. The markers were determined by 2D Hough voting on each slice in the *z*-stack basis (Fig. 5). Hough voting results in a volume with a unique marker on the spheroid, which resembles a column tracking the centre of the nucleus through the focus (see Methods for details). The result of the watershed is a 3D volume with a unique label for each nucleus within the spheroid, which enables the calculation of parameters on individual cells. To compare the mass, volume and mass density distributions, we computed the relative spread,  $\sigma/\mu$ , where  $\sigma$  is the standard deviation and  $\mu$  is the mean associated with the

best Gaussian fit. Our data indicate that the nuclear density (Fig. 5d,  $\sigma/\mu = 0.2$ ) has a much narrower distribution than those of nuclear mass (Fig. 5b,  $\sigma/\mu = 0.9$ ) and volume (Fig. 5c,  $\sigma/\mu = 0.8$ ). These observations indicate that the dry mass density is a much more uniform parameter across different cells. Given the broad distribution of volumes and masses, this result shows that a change in volume is accompanied by an almost linear change in mass.

## Discussion

The principles of AI and confocal microscopy were both formulated in the mid-1950s. Since then, the two technologies have taken independent trajectories, with confocal leading to an entirely new class of scanning imaging modalities and AI giving rise to a variety of applications, from digital assistants to autonomous vehicles. Furthermore, in the past several years, it has become apparent that AI algorithms are valuable tools for extracting knowledge from optical images. As such, the two fields are intersecting again, and this combination seems to hold exciting prospects for biomedicine.

We developed ACM to combine the benefits of non-destructive imaging from QPI with the depth sectioning and chemical specificity associated with confocal fluorescence microscopy. Augmenting an existing laser scanning microscope with a QPI module (LS-GLIM) we can easily collect pairs of registered images from the phase (input data) and fluorescence (ground truth) channels, which allow us to generate co-localized ground truth–input pairs of images. As expected, the transmission quantitative phase image exhibits a much stronger elongation along the *z*-axis, as the scattering wavevector (or momentum transfer) has a much shorter *z*-component than in the reflection

geometry. These pairs of images are used to train a neural network (Efficient U-Net) to perform image-to-image translation from the LS-GLIM to the confocal fluorescence signal. The final ACM image presents the characteristics of the confocal image, with good axial sectioning and chemical specificity (see Figs. 2 and 3). Applying ACM to unlabelled cells allows us to non-destructively translate the confocal microscopy features to dynamic imaging (see Supplementary Videos 1 and 2). In this work we describe the image formation for weak-scattering samples in LS-GLIM. The theoretical model agrees with the experimental data on the system's transfer function by imaging a phase edge.

Confocal geometry advances the QPI imaging capability in several ways, as follows. First, by illuminating one point at a time, the confocal geometry eliminates the spatial cross-talk that affects wide-field methods. With this illumination, the noise from the neighbouring points is lowered. Second, the PMT array provides a much more sensitive detection which, together with the first point above, yields overall higher sensitivity, both spatially and temporally, which we capture in our data. Third, the backscattering geometry yields higher axial spatial frequency coverage, resulting in stronger sectioning. This quality is obvious in our data, both ground truth and inference.

By overcoming the spatial cross-talk limitations associated with wide-field methods, ACM has the potential to provide new data for studying turbid cellular systems. Measuring quantitatively functional parameters from organoids and spheroids can be useful in a variety of applications of biological and clinical relevance. Using the artificial fluorescence images generated by the neural network, we segmented individual nuclei within the 3D structures, which can be used not just for cell counting but also for computing individual nuclear volumes. Furthermore, by creating annotations from the ACM images and applying them back to the input phase images, we extracted dry mass information from individual nuclei, independently from the nuclear volume. Our results indicate that, on average, 42% of the spheroid mass is contained in the nuclei. We also found that the nuclear dry mass density distribution is much narrower than the volume and mass counterparts. Nevertheless, ACM images can differ from confocal fluorescence images. The potential source of error could come from the training corpus including the confocal fluorescence images and LS-GLIM phase images. Specifically, LS-GLIM phase images lack sensitivity to spatial frequencies within the missing cone region, while the confocal fluorescence images are affected by variations in the staining level. Moreover, the contrast in LS-GLIM images comes from the intrinsic inhomogeneity of refractive index distributions in tissues, resulting in much more details compared to the confocal fluorescence counterparts.

Finally, we demonstrated that the network training can be transferred between spheroids suspended in different media with no additional training, which provides versatility to our technique. However, the accuracy of prediction is lower than those of spheroids in PBS, which is expected since the shape of the spheroids in PBS and hydrogel is quite different, and the network is never trained on those data. We anticipate that ACM can be potentially adopted at a broad scale because the LS-GLIM module can be readily added to any existing laser confocal system, while the data for training can be acquired with ease. ACM provides complementary information to that from other laser scanning techniques, as the acquisition is not limited by photobleaching and toxicity, while the axial resolution is maintained at confocal levels.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41566-022-01140-6>.

## References

- Freedman, B. S. et al. Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat. Commun.* **6**, 1–13 (2015).
- Langhans, S. A. Three-dimensional in vitro cell culture models in drug discovery and drug repositioning. *Front. Pharmacol.* **9**, 6 (2018).
- Bissell, M. J., Hall, H. G. & Parry, G. How does the extracellular matrix direct gene expression? *J. Theor. Biol.* **99**, 31–68 (1982).
- Laschke, M. W. & Menger, M. D. Life is 3D: boosting spheroid function for tissue engineering. *Trends Biotechnol.* **35**, 133–144 (2017).
- Bell, C. C. et al. Comparison of hepatic 2D sandwich cultures and 3D spheroids for long-term toxicity applications: a multicenter study. *Toxicol. Sci.* **162**, 655–666 (2018).
- Fong, E. L. S., Toh, T. B., Yu, H. & Chow, E. K.-H. 3D culture as a clinically relevant model for personalized medicine. *SLAS Technol.* **22**, 245–253 (2017).
- Kamm, R. D. et al. Perspective: the promise of multi-cellular engineered living systems. *APL Bioeng.* **2**, 040901 (2018).
- Cvetkovic, C. et al. Three-dimensionally printed biological machines powered by skeletal muscle. *Proc. Natl Acad. Sci. USA* **111**, 10125–10130 (2014).
- Williams, B. J., Anand, S. V., Rajagopalan, J. & Saif, M. T. A. A self-propelled biohybrid swimmer at low Reynolds number. *Nat. Commun.* **5**, 1–8 (2014).
- Chen, X. & Korotkova, O. Optical beam propagation in soft anisotropic biological tissues. *Osa Continuum* **1**, 1055–1067 (2018).
- Tuchin, V. V. & Society of Photo-optical Instrumentation Engineers. *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis* 2nd edn (SPIE/International Society for Optical Engineering, 2007).
- Chen, W. et al. High-throughput image analysis of tumor spheroids: a user-friendly software application to measure the size of spheroids automatically and accurately. *J Vis Exp.* <https://doi.org/10.3791/51639> (2014).
- Minsky, M. S. Memoir on inventing the confocal. *Scanning Microsc.* **10**, 128–138 (1988).
- Wilson, T. & Sheppard, C. *Theory and Practice of Scanning Optical Microscopy* (Academic, 1984).
- Diaspro, A. *Optical Fluorescence Microscopy* (Springer, 2011).
- Lippincott-Schwartz, J., Altan-Bonnet, N. & Patterson, G. H. Photobleaching and photoactivation: following protein dynamics in living cells. *Nat. Cell Biol.* **S7–S14** (2003).
- Hoebe, R. A. et al. Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nat. Biotechnol.* **25**, 249–253 (2007).
- Crivat, G. & Taraska, J. W. Imaging proteins inside cells with fluorescent tags. *Trends Biotechnol.* **30**, 8–16 (2012).
- Graf, B. W. & Boppart, S. A. in *Live Cell Imaging: Methods and Protocols* 211–227 (Springer, 2010).
- North, A. J. Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. *J. Cell Biol.* **172**, 9–18 (2006).
- Hoover, E. E. & Squier, J. A. Advances in multiphoton microscopy technology. *Nat. Photon.* **7**, 93–101 (2013).
- Stelzer, E. H. et al. Light sheet fluorescence microscopy. *Nat. Rev. Methods Primers* **1**, 1–25 (2021).
- Choi, W. J., Pepple, K. L. & Wang, R. K. Automated three-dimensional cell counting method for grading uveitis of rodent eye in vivo with optical coherence tomography. *J. Biophoton.* **11**, e201800140 (2018).
- Huang, Y. et al. Optical coherence tomography detects necrotic regions and volumetrically quantifies multicellular tumor spheroids. *Cancer Res.* **77**, 6011–6020 (2017).



25. Schnell, M. et al. High-resolution label-free imaging of tissue morphology with confocal phase microscopy. *Optica* **7**, 1173–1180 (2020).
26. Hase, E. et al. Scan-less confocal phase imaging based on dual-comb microscopy. *Optica* **5**, 634–643 (2018).
27. Singh, V. R. et al. Studying nucleic envelope and plasma membrane mechanics of eukaryotic cells using confocal reflectance interferometric microscopy. *Nat. Commun.* **10**, 1–8 (2019).
28. Liu, C. et al. High-speed line-field confocal holographic microscope for quantitative phase imaging. *Opt. Express* **24**, 9251–9265 (2016).
29. Popescu, G. *Quantitative Phase Imaging of Cells and Tissues* (McGraw-Hill, 2011).
30. Park, Y., Depeursinge, C. & Popescu, G. Quantitative phase imaging in biomedicine. *Nat. Photon.* **12**, 578 (2018).
31. Chen, X., Kandel, M. E. & Popescu, G. Spatial light interference microscopy: principle and applications to biomedicine. *Adv. Opt. Photon.* **13**, 353–425 (2021).
32. Chen, X., Kandel, M. E., Hu, C., Lee, Y. J. & Popescu, G. Wolf phase tomography (WPT) of transparent structures using partially coherent illumination. *Light Sci. Appl.* **9**, 1–9 (2020).
33. Merola, F. et al. Tomographic flow cytometry by digital holography. *Light Sci. Appl.* **6**, e16241 (2017).
34. Chen, M., Ren, D., Liu, H.-Y., Chowdhury, S. & Waller, L. Multi-layer Born multiple-scattering model for 3D phase microscopy. *Optica* **7**, 394–403 (2020).
35. Hugonnet, H. et al. Multiscale label-free volumetric holographic histopathology of thick-tissue slides with subcellular resolution. *Adv. Photon.* **3**, 026004 (2021).
36. Ledwig, P. & Robles, F. E. Epi-mode tomographic quantitative phase imaging in thick scattering samples. *Biomed. Opt. Express* **10**, 3605–3621 (2019).
37. Kandel, M. E. et al. Epi-illumination gradient light interference microscopy for imaging opaque structures. *Nat. Commun.* **10**, 4691 (2019).
38. Nguyen, T. H., Kandel, M. E., Rubessa, M., Wheeler, M. B. & Popescu, G. Gradient light interference microscopy for 3D imaging of unlabeled specimens. *Nat. Commun.* **8**, 210 (2017).
39. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
40. Wu, Y. et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat. Methods* **16**, 1323–1331 (2019).
41. Williams, B. M. et al. An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study. *Diabetologia* **63**, 419–430 (2020).
42. Cheng, S., Li, H., Luo, Y., Zheng, Y. & Lai, P. Artificial intelligence-assisted light control and computational imaging through scattering media. *J. Innov. Opt. Health Sci.* **12**, 1930006 (2019).
43. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
44. Lee, M. et al. Deep-learning-based three-dimensional label-free tracking and analysis of immunological synapses of CAR-T cells. *eLife* **9**, e49023 (2020).
45. Tan, M. & Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proc. 36th International Conference on Machine Learning* 6105–6114 (PMLR, 2019).
46. Chen, X. & Korotkova, O. Probability density functions of instantaneous Stokes parameters on weak scattering. *Opt. Commun.* **400**, 1–8 (2017).
47. Kanai, Y. & Hirokawa, N. Sorting mechanisms of Tau and MAP2 in neurons: suppressed axonal transit of MAP2 and locally regulated microtubule binding. *Neuron* **14**, 421–432 (1995).
48. Rein, A. Retroviral RNA packaging: a review. *Arch. Virol. Suppl.* **9**, 513–522 (1994).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023

<sup>1</sup>Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>2</sup>Department of Computer Science and Engineering, Washington University in St Louis, St Louis, Missouri, USA. <sup>3</sup>Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>4</sup>Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>5</sup>Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>6</sup>Department of Molecular and Integrative Physiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>7</sup>Carl Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>8</sup>Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>9</sup>Present address: School of Applied and Engineering Physics, Cornell University, Ithaca, USA. <sup>10</sup>Present address: Groq, Mountain View, CA, USA. ✉ e-mail: [xc289@cornell.edu](mailto:xc289@cornell.edu)

## Methods

### ACM system

The experimental set-up for ACM is a multichannel imaging system, which consists of confocal microscopy (LSM 900, Zeiss) and LS-GLIM. The LS-GLIM module upgrades a laser scanning confocal microscope outfitted with DIC optics by providing phase-shifting assembly capability (Fig. 1a). The laser scanning interference microscope shares the same two-laser lines (488 nm, 561 nm) of the confocal microscope. The laser source from the confocal microscopy goes up through the matched DIC prism and objective ( $\times 63$ ,  $\times 40$ ) and then is scattered by the sample. After the sample, the light is collected by the condenser of the DIC microscope. The light then travels through the phase-shifting assembly, which consists of an LCVR (Thorlabs) followed by a linear polarizer. We removed the analyser that normally sits inside the condenser to allow the liquid crystal to modulate the phase shift between the two orthogonal polarizations. The stabilization time of the LCVR is approximately 70 ms. Four intensity frames are recorded by the photomultiplier tube (PMT, Zeiss) corresponding to each  $\pi/2$  phase shift, as shown in Fig. 1b. The acquisition time of each frame is approximately the same as for a confocal fluorescence image, which depends on the dwell time and pixel numbers set for the image acquisition. The dwell time for all the images was chosen to be 1.2  $\mu$ s, such that the acquisition time is  $\sim 3.7$  s for an image with  $1,744 \times 1,744$  pixels. The quantitative phase images are generated in real-time by the phase-retrieval reconstruction algorithm and Hilbert transform algorithm<sup>38</sup>. The system registers pairs of z-stack images from both the confocal fluorescence and quantitative phase, which serve, respectively, as ground truth and input images for machine learning (Fig. 1b). The z-sampling was chosen to be 0.2  $\mu$ m, 0.2  $\mu$ m and 1  $\mu$ m for microbeads, neurons and spheroids, respectively. The x-y sampling was 0.09  $\mu$ m for all of the data presented in this paper.

### Network training

We trained E-U-Nets with paired phase and fluorescent images. The input channels of an E-U-Net are three neighbouring phase slices, and the output is the corresponding central fluorescent slice. This network design allows an E-U-Net to use information from phase images acquired at multiple neighbouring imaging planes to better predict the fluorescent image.

The network architecture of a multichannel E-U-Net is shown in Fig. 2a and Supplementary Fig. 7. It represents a modification of a standard U-Net where the encoder is replaced with an EfficientNet<sup>45</sup>. The EfficientNet generally has a powerful capacity for feature extraction but is relatively small in network size. Training an E-U-Net from scratch can be challenging when the number of paired phase and fluorescent images is limited. A transfer learning strategy was used in the E-U-Net training to mitigate this challenge. Specifically, the weights of the EfficientNet encoder were initialized with weights pre-trained on an ImageNet dataset<sup>49</sup> for an image classification task. The ImageNet is a benchmark image set that contains millions of labelled nature images.

In this study, a neuron dataset, a spheroid cell dataset and a bead dataset were used for training, validating and testing the E-U-Nets, respectively. The neuron dataset contained 22 image stacks that each contained 300 neuron phase images of size  $1,744 \times 1,744$  pixels and their related two-channel fluorescent images, which correspond to fluorescent signals from Tau and MAP2 proteins, respectively. The spheroid cell dataset contained 21 stacks that each contained 100 spheroid cell phase images of size  $1,744 \times 1,744$  pixels and the related two-channel fluorescent images, which correspond to fluorescent signals from DNA and RNA, respectively. The bead dataset contained eighteen image stacks that each contained 250 bead phase images of size  $128 \times 128$  pixels and the associated fluorescent images. To facilitate network training, the pixel values in each fluorescent image stack were scaled to a range of [0, 255.0]. This was accomplished as:  $x_o = 255.0 \times \frac{x_i - x_{0.01\%}}{x_{99.99\%} - x_{0.01\%}}$ , where  $x_{0.01\%}$  and  $x_{99.99\%}$  represent the 0.01%th and 99.99%th values among all the pixel values in the image stack after they

were sorted in non-decreasing order;  $x_i$  and  $x_o$  represent the original and scaled value of a pixel, respectively. The estimated fluorescent image stack was subsequently rescaled to its original range using  $x_i = \frac{x_o}{255.0} (x_{99.99\%} - x_{0.01\%}) + x_{0.01\%}$ . For those image stacks without ground truth values, the  $x_{0.01\%}$  and  $x_{99.99\%}$  can be estimated as the average of  $x_{0.01\%}$  and  $x_{99.99\%}$  related to the ground truth values in the training set.

Considering the limited number of image stacks in the three datasets described above, a threefold cross-validation approach was employed to train and validate the E-U-Nets after a few testing image stacks were held out for E-U-Net testing. For a given dataset in which the testing stacks have been held out, the threefold cross-validation approach involves randomly dividing all the stacks in the dataset into three folds of approximately equal size. The first two folds and the remaining one-fold were treated as a training set and a validation set to train and validate E-U-Nets, respectively. The procedure was repeated three times; each time, a different fold was treated as the validation set. The three procedures resulted in the validation of the E-U-Nets on each image stack. The trained E-U-Nets were finally tested on the held out unseen testing samples. Details related to the cross-validation of E-U-Nets on the neuron, spheroids cell and bead datasets are described below.

For the neuron dataset, two separate E-U-Nets were trained: one to translate phase images into each of the two-channel fluorescent images. The EfficientNet-B7 network was employed in the two E-U-Nets. The network architecture of the EfficientNet-B7 is shown in Supplementary Fig. 7. Two neuron image stacks were held out as unseen testing data; the remaining twenty stacks were employed in the threefold cross-validation process described above. In the threefold cross-validation process, the twenty image stacks were randomly divided into three folds that contained six, seven and seven image stacks, respectively. For each data split, the E-U-Nets were trained by minimizing a mean square error (MSE) loss function that measures the difference between the predicted fluorescent images and their corresponding ground truth values. The loss function was minimized by the use of an ADAM optimizer<sup>50</sup> with a learning rate of  $5 \times 10^{-4}$ , which was empirically determined. In each training iteration, a batch of paired three neighbouring phase images and the corresponding central fluorescent image were sampled from the training image stacks and then randomly cropped into patches of  $515 \times 512$  pixels as training samples to train the networks. The batch size was set to four. A decaying strategy was applied to the learning rate to mitigate the overfitting by multiplying the learning rate by 0.8 when the validation MSE loss did not decrease for consecutive epochs. An epoch is a sequence of iterations that walk through all the image slices in the training set. The validation MSE loss was computed between the predicted fluorescent images and their ground truth values for validation images. In the network training, an early stopping strategy was employed to determine the end of the network training. Specifically, at the end of each epoch, the being-trained E-U-Net model was evaluated by computing the average of the PCCs between the predicted fluorescent images and the related ground truth values. The network training stopped if the average validation PCC did not increase for twenty epochs as shown in Supplementary Fig. 8. The two figures show the average training and validation stopping rule metric for training the two E-U-Nets respectively in one of the three training procedures of the threefold cross-validation process. After the E-U-Nets were trained, the performances of the trained networks were evaluated on the validation set by computing the PSNR and PCC between the predicted fluorescent stacks and the related ground truth values. The threefold cross-validation process resulted in validation results for each of the twenty stacks. These validation results were combined and are reported in Supplementary Note 4. The E-U-Nets trained in the cross-validation process were also tested in the two unseen stacks. The corresponding PCCs and PSNRs are presented in Supplementary Note 4.

For the spheroids cell dataset, two separate E-U-Nets were trained for each fluorescent channel. The EfficientNet-B7 network was employed as the encoder in the two trained E-U-Nets. Two spheroid cell image stacks were held out for E-U-Net testing; the remaining nineteen stacks were randomly split into three folds that contain six, six and seven stacks, respectively, in the three-fold cross-validation process. The other training settings were the same as those described above for network training on the neuron dataset. The training and validation PCCs over epochs correspond to training the two E-U-Nets in one of the three training procedures of the three-fold cross-validation process are displayed in Supplementary Fig. 9. The threefold cross-validation results related to PSNRs and PCCs are reported in Supplementary Note 4. The results tested on two unseen testing stacks are also shown in Supplementary Note 4.

For the bead dataset, a single E-U-Net was built for the phase-to-fluorescent image translation. EfficientNet-B0 was employed as the encoder in the E-U-Net. The architecture of the EfficientNet-B0 network is shown in Supplementary Fig. 7. One of the bead image stacks was held out as an unseen testing stack for the E-U-Net testing; the remaining seventeen bead stacks were randomly divided into three folds that each contains five, six and six image stacks, respectively, for the threefold cross-validation process. Paired images of size  $128 \times 128$  pixels were employed for the E-U-Net training. The batch size was 32. The other training settings were the same as those for the network training on neuron and spheroid cell datasets, as described above. The training and validation stopping rule metric over epochs for one of the three training procedures of the three-fold cross-validation process are displayed in Supplementary Fig. 10. The threefold cross-validation results related to PSNR and PCC performances are reported in Supplementary Note 4. The results on the unseen bead stack are also shown in Supplementary Note 4.

The E-U-Nets were implemented by use of the Python programming language with libraries including Python 3.6 and TensorFlow 1.14. Model training, validation and testing were performed on an NVIDIA Tesla V100-GPU with 32 GB VRAM. E-U-Net training on the neuron dataset and spheroid dataset took approximately 24 h. E-U-Net training on the bead dataset took approximately 2 h. The inference time for a fluorescent image slice of  $1,744 \times 1,744$  pixels was approximately 400 ms.

### Neuron analysis

The volume of neurons was calculated from the ACM images using binary masks with background thresholding. The 3D dry mass distribution was generated with the multiplication of binary masks and the 3D dry mass distribution from the QPI images<sup>51</sup>. The 3D dry mass density is linearly related to the depth-resolved phase maps as

$$M(x, y, z) = \frac{\lambda}{2\pi\gamma\delta z} \phi(x, y, z), \quad (2)$$

where  $\lambda$  is the wavelength of the illumination and the refractive increment  $\gamma \approx 0.2$ , which lies within the  $0.18\text{--}0.21 \text{ ml g}^{-1}$  range for most biological samples<sup>52</sup>;  $\delta z$  represents  $z$ -sampling, which is  $-1 \mu\text{m}$  for our LS-GLIM;  $\phi(x, y, z)$  is the measured phase image on each  $z$ -plane.

### Spheroid analysis

Three-dimensional semantic segmentation maps were generated from the estimated fluorescent signals corresponding to the RNaselect and 7-ADD stains by applying fixed thresholds for the entire data. This map of RNA- and DNA-stained regions was further refined by assigning a nucleoli label to the RNA inside of the DNA regions. To generate a map labelling the spheroid, a threshold was applied to the quantitative phase signal after Hilbert demodulation<sup>51</sup>. Fields of view were acquired to contain a single spheroid, and phase values coincident with the assigned label (nucleus, spheroid) were tallied on a per-spheroid basis to report on the dry mass and volume.

### Automated 3D cell counting

To segment our images into individual nuclei, we used a 3D variation of the marker-controlled watershed on the estimated DNA images<sup>53</sup>. We note that the ACM data lacked the unwanted pixel-level noise typically associated with photon-starved fluorescent images. This technique requires the image to be annotated into sample and background regions with a non-overlapping marker used to identify the cell. We performed 2D Hough voting which is used to identify the centre of the nucleus in each  $z$ -slice, producing what resembles a curve through the  $z$ -dimension. To regularize our approach, we applied a  $3 \times 3$  blur to correct for minor disconnects in our segmentation algorithm. The result of our watershed approach is a 3D volume with a unique label annotating each nucleus (Fig. 5a). To validate our method, we compared our results to a manual cell count performed in AMIRA (version 5.4.3)<sup>54</sup>. We obtained 142 cells counted automatically versus 136 cells counted manually (4% error). The principal disagreement was due to undercounting touching cells. This procedure was implemented in MATLAB using the `imfindcircles` and `watershed` commands.

### Sample preparation

**Hippocampal neuron preparation.** All procedures involving animals were reviewed and approved by the Institutional Animal Care and Use Committee at the University of Illinois Urbana-Champaign and conducted per the guidelines of the US National Institute of Health. For our neuron imaging experiments, we used primary hippocampal neurons harvested from dissected hippocampi of Sprague–Dawley rat embryos at embryonic day 18. Dissociated hippocampal neurons were plated on multiwell plates (Cellvis, P06-20-1.5-N) that were pre-coated with poly-D-lysine ( $0.1 \text{ mg ml}^{-1}$ ; Sigma-Aldrich). Hippocampal neurons were incubated for 3 h at  $37^\circ\text{C}$  and under 5%  $\text{CO}_2$  in a plating medium containing 86.55% Eagle's MEM with Earle's BSS (Lonza), 10% foetal bovine Serum (refiltered, heat-inactivated; ThermoFisher), 0.45% of 20% (wt/vol) glucose, 1 equiv. 100 mM sodium pyruvate (100x; Sigma-Aldrich), 1 equiv. 200 mM glutamine (100x; Sigma-Aldrich) and 1 equiv. penicillin/streptomycin (100x; Sigma-Aldrich) to help attachment of neurons ( $300 \text{ cells per mm}^2$ ). The plating media was aspirated and replaced with maintenance media containing Neurobasal growth medium supplemented with B-27 (Invitrogen), 1% 200 mM glutamine (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) and incubated for 10 days at  $37^\circ\text{C}$ , in the presence of 5%  $\text{CO}_2$ . Hippocampal neurons were maintained for 2 weeks before performing immunostaining.

### Immunostaining protocol

Neurons were stained with antibodies for Tau (Abcam, ab80579) and MAP2 (Abcam ab32454) to localize axons and dendrites. Neurons were fixed with freshly prepared 4% paraformaldehyde for 15 min following 0.5% Triton-X for 10 min and 2% bovine serum albumin (BSA, ThermoFisher) for 2 h incubation in  $4^\circ\text{C}$ . Hippocampal neurons were incubated for 8 h at  $4^\circ\text{C}$  with anti-Tau antibodies that were diluted to 1:250 in 5% BSA. After washing with PBS, neurons were exposed for 8 h at  $4^\circ\text{C}$  to goat anti-mouse secondary antibody (Abcam, ab205719) which was diluted to 1:500 in 5% BSA. Hippocampal neurons were then incubated in anti-MAP2 antibody (1:500 dilution) in 5% BSA for 8 h, followed by goat antirabbit secondary antibody (Abcam, ab205718, 1:1000 dilution) in 5% BSA for 8 h at  $4^\circ\text{C}$ .

### Liver cancer spheroid (HepG2 cells)

Human hepatocarcinoma cells (HepG2, ATCC) were cultured in T-75 flasks with DMEM (Thermo), 10% foetal bovine serum and 1% penicillin-streptomycin (Gibco) for 7 days, leading to spontaneous pre-formed spheroids. The flasks were incubated at  $37^\circ\text{C}$  and 5%  $\text{CO}_2$ . The media were replaced every two to three days. Spheroids were incubated with TrypLE Express (Thermo) for 10 min to detach pre-formed spheroids of approximately  $100\text{--}200 \mu\text{m}$  in diameter from the culture flask. The passage number used was between two and six.

Pre-formed spheroids were plated on poly-D-lysine coated glass-bottom dishes. The spheroids were incubated for 10 min to allow for attachment. They were then covered with a collagen hydrogel (bovine collagen type 1, Advanced Biomatrix). The cells were incubated for three days to allow for cellular reorganization into a regular spheroidal shape. The spheroids were first fixed in a 1:1 ratio of methanol: acetone at  $-4^{\circ}\text{C}$  for 20 min. Cells fixed using this method do not need an additional permeabilization step due to the acetone. The cell nucleus was stained using 7-AAD (red, 6163, ThermoFisher) by adding  $1\ \mu\text{l}$  of the stock stain into 1 ml of PBS. The cell RNA was stained using SYTO RNASelect Green (S32703, ThermoFisher) by first creating a  $5\ \mu\text{M}$  working solution and then adding  $100\ \mu\text{l}$  of the working solution to  $900\ \mu\text{l}$  of PBS. The samples were stained at room temperature for 30 min before rinsing once. Two types of samples in PBS or hydrogel were imaged after staining.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Due to size considerations, the data that support the findings of this study are available from the corresponding author on reasonable request.

### Code availability

The code that supports the findings of this study are available from the corresponding author on reasonable request.

### References

49. Kingma, D.P. & Ba, J. Adam: a method for stochastic optimization. CoRR, abs/1412.6980 (2014).
50. Deng, J. et al. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009).
51. Kandel, M. E. et al. Multiscale assay of unlabeled neurite dynamics using phase imaging with computational specificity. *ACS Sens.* **6**, 1864–1874 (2021).
52. Barer, R. Determination of dry mass, thickness, solid and water concentration in living cells. *Nature* **172**, 1097–1098 (1953).
53. Gonzalez, R. C. & Woods, R. E. *Digital Image Processing* (Prentice-Hall, 2002).
54. Stalling, D., Westerhoff M., & Hege H.-C. Amira: A highly interactive system for visual data analysis. *Visualization Handb.* **38**, 749–767 (2005).

### Acknowledgements

This work is supported by the National Science Foundation (grant nos. CBET0939511 STC, NRT-UtB 1735252, CBET-1932192), the National Institute of General Medical Sciences (grant no. GM129709), the National Institute of Neurological Disorders and Stroke (grant nos. NS097610 and NS100019) and the National Cancer Institute (grant no. CA238191).

### Author contributions

X.C., M.E.K., and G.P. conceived the project. X.C. and M.E.K. designed the experiments. X.C. and M.E.K. built the system. X.C. performed imaging. S.H. trained the machine learning network. X.C. and M.E.K. analysed the data. G.T & H.J.C. provided neurons. Y.J.L. cultured neurons and performed immunocytochemistry. K.M.S. & H.K. provided spheroids. X.C., C.H. and G.P. derived the theoretical model. X.C., M.E.K., S.H., C.H. and G.P. wrote the manuscript. M.A. supervised the AI work. G.P. supervised the project.

### Competing interests

G.P. had a financial interest in Phi Optics, a company developing QPI technology for materials and life science applications. The remaining authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41566-022-01140-6>.

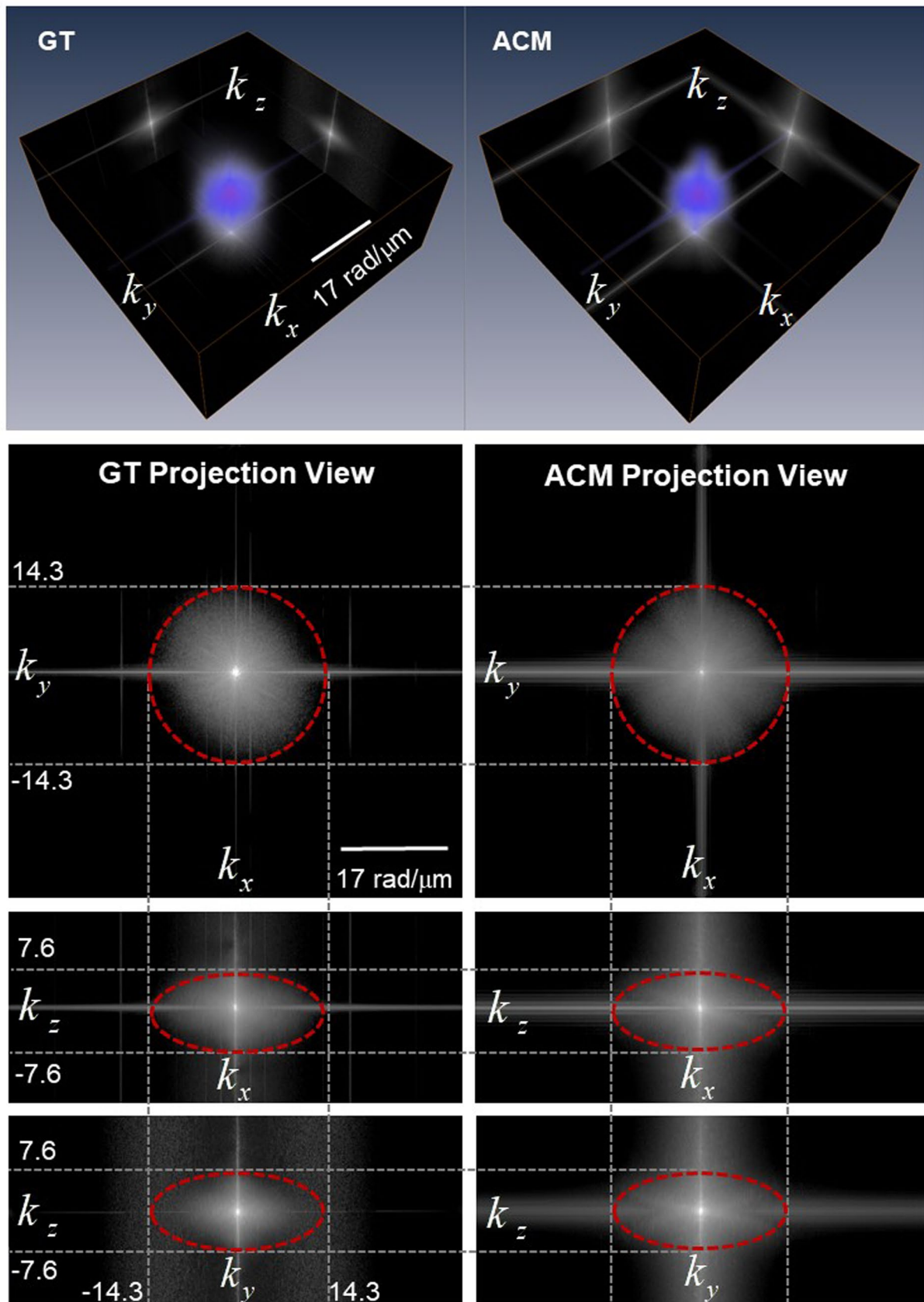
**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41566-022-01140-6>.

**Correspondence and requests for materials** should be addressed to Xi Chen.

**Peer review information** *Nature Photonics* thanks Adam Wax and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Comparison of GT to ACM power spectra from Fig. 3 a-l



Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Comparison of ground truth to ACM power spectra from Fig. 3a–f.** Contours circumscribing theoretical resolution limits of confocal fluorescence system (ground truth) are shown in as red dotted circles. The theoretical lateral resolution of the system is  $0.22\ \mu\text{m}$  (NA = 1.3, 1 Airy Unit (AU), excitation wavelength at 561 nm), corresponding to a maximum lateral

frequency of  $14.3\ \text{rad}/\mu\text{m}$ . The theoretical axial resolution of the system is about  $0.50\ \mu\text{m}$ , corresponding to a maximum axial frequency of  $6.3\ \text{rad}/\mu\text{m}$ . The 3D frequency coverage of the ground truth and ACM spectra agree, and both reach the theoretical resolution limits.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

### Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

### Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

### Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

### Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size      Sample sizes were determined by the available resources. All available samples were used in this study and found to be statistically significant.

Data exclusions      No data was excluded.

Replication      Replication was performed by imaging multiple sample types as shown in the study. All attempts of replication were successful.

Randomization      Data were randomly split into training, validation, and test data sets.

Blinding      The data collection and analysis were performed by several individuals at different independent phases in this study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a      Involved in the study

Antibodies

Eukaryotic cell lines

Palaeontology and archaeology

Animals and other organisms

Clinical data

Dual use research of concern

### Methods

n/a      Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging



## Antibodies

---

Antibodies used

Tau (Abcam, ab80579) and MAP2 (Abcam ab32454)

Validation

Popular antibodies validated by the manufacturer

## Eukaryotic cell lines

---

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HepG2 (ATCC), hippocampal neurons (Sprague-Dawley rats, primary)

Authentication

none

Mycoplasma contamination

no testing or relevance to imaging experiments

Commonly misidentified lines  
(See [ICLAC](#) register)

*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*