



Multicolor localization microscopy and point-spread-function engineering by deep learning

ERAN HERSHKO,^{1,2,3} LUCIEN E. WEISS,^{2,3} TOMER MICHAELI,¹ AND YOAV SHECHTMAN,^{2,*}

¹Electrical Engineering Department, Technion, 32000 Haifa, Israel

²Biomedical Engineering Department, Technion, 32000 Haifa, Israel

³Equal contribution

*yoavsh@bm.technion.ac.il

Abstract: Deep learning has become an extremely effective tool for image classification and image restoration problems. Here, we apply deep learning to microscopy and demonstrate how neural networks can exploit the chromatic dependence of the point-spread function to classify the colors of single emitters imaged on a grayscale camera. While existing localization microscopy methods for spectral classification require additional optical elements in the emission path, *e.g.*, spectral filters, prisms, or phase masks, our neural net correctly identifies static and mobile emitters with high efficiency using a standard, unmodified single-channel configuration. Furthermore, we show how deep learning can be used to design new phase-modulating elements that, when implemented into the imaging path, result in further improved color differentiation between species, including simultaneously differentiating four species in a single image.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Single-particle tracking and super-resolution fluorescence microscopy harness a high signal-to-background ratio to attain nanoscale spatial information exceeding the diffraction limit. Localization-based microscopy techniques: (F)PALM, STORM [1–3], and related methods [4–8], can improve the resolution of an image by an order of magnitude relative to that attained in normal epifluorescence microscopy, routinely attaining 10–40 nm spatial resolution. The key idea of localization microscopy is to find the likely underlying position of an emitter whose emission produces a diffraction-limited spot captured by a camera (*e.g.* the center of the point-spread function, PSF). It has been shown previously, that in addition to the x-y position, other information can be extracted from the standard PSF as well, such as photophysical properties, emitter depth, orientation, and directionality of motion [9–13].

Of particular importance for biological imaging is real-time, correlative information between multiple species in a sample [14] (*e.g.* proteins or organelles in cells and tissues). Typically, this is achieved by attaching spectrally-distinctive fluorophores to molecules of interest, thus necessitating multicolor imaging. Using an RGB camera is not practical for low-signal applications where photons are precious, such as single-particle tracking and single-molecule microscopy. The popular approaches for multicolor microscopy either divide the emission spectra between multiple sensors [15,16], or sequentially image one species at a time by switching spectral filters and/or light sources [14]. The former method requires precise registration between the channels [17], while the latter is limited to imaging quasi-static objects and is prone to artifacts caused by non-simultaneous, slower acquisitions, *e.g.* sample drift, cell growth and migration.

Recently, multicolor localization microscopy by PSF engineering has been demonstrated [18,19]. In this technique, the image that each point source creates on the camera, namely, the PSF, is modified to encode the color of the emitter. In other words – molecules emitting different colors produce images of different shapes. This modification is performed by adding

a spectrally-sensitive, phase-modulating element, *e.g.* a liquid-crystal spatial light modulator (SLM) to the imaging path which is positioned in a plane conjugate to the back focal plane of the microscope objective [20]. The main advantage of PSF engineering is that it enables truly simultaneous imaging and tracking of multiple, colored emitters with no compromise in the field-of-view (FOV). However, the design of a multicolor PSF that optimally balances between emitter detectability, which requires the PSF to be small, and color classification, which requires the PSF to vary significantly as a function of wavelength, is still an open question.

In recent years, deep learning has been employed with great success in a variety of tasks [21], including designing optical systems [22–24] and interpreting single-molecule data to produce super-resolution images [25–27]. The multi-layer architecture of neural nets allows for extraction of complex features from data, while distilling the desired information from an input. Neural nets are capable of learning to recognize subtle features, even under adverse imaging conditions, making the technique well suited for the problem of species differentiation in localization microscopy where the PSF differs only slightly between different wavelengths.

Here, we present two fundamental contributions by applying deep learning to microscopy. First, we experimentally demonstrate an algorithm for determining an emitter's color using a standard fluorescence microscope equipped with a grayscale camera with no additional hardware modification (Fig. 1). This is enabled by the fact that the PSF of any optical system is dependent on the wavelength, even without PSF engineering. Second, we develop and experimentally demonstrate an additional neural net that algorithmically generates an optimized, color-encoding PSF using phase modulation, for maximal color-distinguishability when there are several differently-colored emitters present.

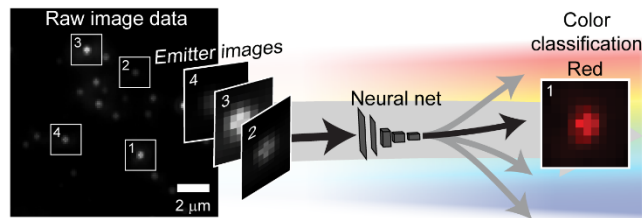


Fig. 1. Color classification with neural nets. Patches containing red and green Qdots are extracted from a grayscale image and classified by a neural net. An example of a red emitter classification is depicted.

2. Results

The PSF of any imaging system depends on the emitter's wavelength, due to diffraction and chromatic aberrations [28]. To test first whether we could discriminate between two types of emitters imaged on a grayscale camera, we prepared a thin sample containing green and red quantum dots (Qdots) with emission peaks at 565 and 705 nm, and imaged it using an epifluorescence microscope (Fig. 2(a)). For each field of view (FOV), three image sets were recorded: first, a grayscale image containing all of the Qdots (Fig. 2(b)); second, an image with a spectral long-pass filter added so that only the red Qdots were visible (Fig. 2(c), red channel); and lastly, a bandpass filter so that only the green Qdots were visible (Fig. 2(c), green channel). The shapes of the average emitters for each colors were found to be slightly different (Fig. 2(e)), and the brightness distributions of the two types of emitters were highly overlapping (Fig. 2(f)); although, the red beads were ~twice as bright on average with 21,130 and 9,520 mean signal photons detected for the red and green Qdots, respectively, and a mean background standard deviation of 6.5 photons per pixel.

Asymmetric 2D-Gaussian fits performed on each ROI were not readily separable by the fit parameters (Fig. 2(g)) however, a 2D t-Distributed Stochastic Neighbor Embedded (t-

SNE) projection [29] of the pixel values for each emitter-centered ROI suggested that there existed separable differences between the two types of emitters that could be exploited (see Appendix 5.3.1.). Based on this result, we attempted to classify emitters through various algorithms: nearest-neighbor search, matched filters, and a cutoff for emitter intensity and Gaussian-fit shape sizes (Fig. 2(h), Appendix 5.3.2.). These results were compared to that of a deep convolutional neural net trained on twenty FOVs containing ~ 400 Qdots per field (training details are described in Methods, and illustrations are provided in Appendix 5.1.). The net was used to classify Qdots in six new FOVs (Fig. 2(d), [Visualization 1](#)). While the trained net correctly predicted $96.4\% \pm 1.2\%$ of Qdots ($N_{\text{total}} = 2491$, where the percentage is the weighted mean and the uncertainty is represented by the weighted standard deviation, estimated from six FOVs), other methods tested were unable to perform nearly as well; for instance, an optimized nearest-neighbor algorithm based on the pixel-wise values correctly classified only $\sim 76.7\%$ of the Qdots (Fig. 2(h)).

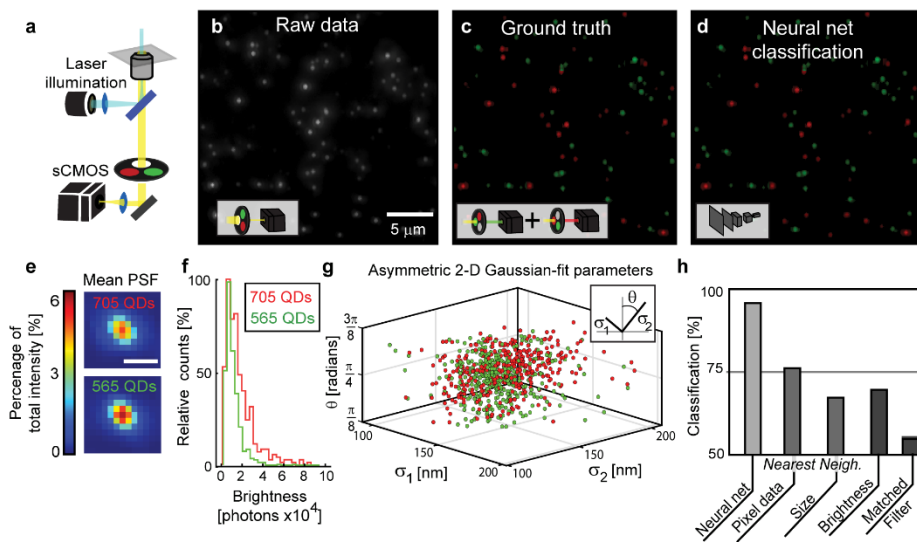


Fig. 2. Qdot color determination using neural nets. (a) An epi-illumination microscope was used to examine Qdots on a glass coverslip. (b) A grayscale image of red and green Qdots. (c) A color image of the same sample obtained by imaging with two spectral filters. (d) The color-classified image from the neural net. (e) Average PSFs for red and green Qdots. (f) A histogram of the signal photons of the two Qdots. (g) A 3D scatter plot of the red and green Qdots showing the fitted parameters from an astigmatic Gaussian with two shape parameters (σ_1 and σ_2) and a variable angle (θ). (h) Classification percentage for emitters by various methods such as Nearest Neighbors, parameter thresholds, and matched filtering.

While classifying two types of emitters using the standard PSF is possible, various other solutions exist for imaging chromatically well-separated emitters, *e.g.* splitting the imaging channels with a single dichroic mirror or switching filters (for stationary samples). A much more challenging problem is simultaneously imaging more types of emitters, where splitting the imaging channels multiple times is more cumbersome and expensive. With the standard PSF, our initial approach showed significantly-reduced effectiveness relative to the two-color classifier, described in further detail later; therefore, we employed PSF engineering, where a wavelength-dependent aberration is added to help unquify each color [18,30,31]. Phase-modulating elements, placed in the back focal plane of the objective, can be designed to engineer PSFs (Fig. 3(a)); but how should one design the optimal PSF to both achieve precise localization of emitters and simultaneously classify their color? Previous work on optimal PSF design for single-color 3D imaging used Fisher information as a design metric [32,33], but it is unclear how to take an analogous approach for multiple wavelengths and spatial

positions. Here, we solve the PSF-design challenge using deep learning – a particularly appealing method in this case, since we can directly optimize emitter detection and color classification at once.

To engineer an optimized PSF for localization and color determination by deep learning, we designed a neural net composed of two parts (Fig. 3(b)). The first portion is a simulator, referred to hereafter as the SLM-optimizer, which generates realistic image data of point sources. These images are fed to the second piece, the Reconstruction net, which outputs the positions and color classifications for detected emitters in the image. The SLM-optimizer and the Reconstruction net both use the difference between the simulated ground truth and the Reconstruction to optimize the SLM voltage mask and the associated Reconstruction-net parameters (See Methods and Visualization 2). Using this approach, we optimized phase masks for 2, 4 and 5-color samples (Appendix 5.2.6.).

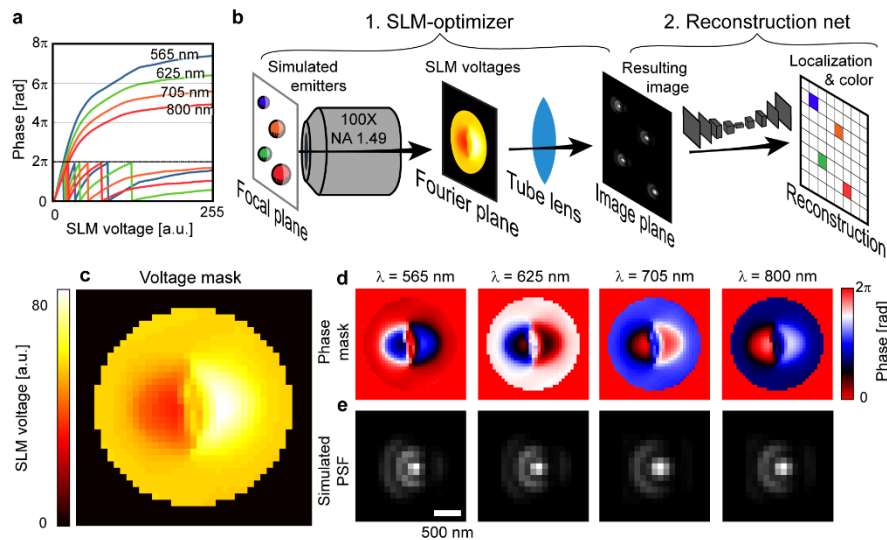


Fig. 3. Design of an optimized SLM pattern using neural networks. (a) An SLM imparts a chromatically-dependent phase delay as a function of applied voltage. (b) A schematic depicting the process for creating an optimized phase mask consisting of 1. an SLM optimizer, used to generate the resulting PSFs for a particular SLM voltage pattern, and 2. a reconstruction net, which decodes the generated images. (c) The optimized SLM voltage pattern for color determination by a neural net. (d) The phase delay imparted to 565, 625, 705 and 800 nm light. (e) Simulated PSFs.

The net converged to a novel SLM-voltage pattern (Fig. 3(c)), which imparts distinct phase-delay patterns depending on the wavelength (Fig. 3(d)), yielding different PSFs for each color (Fig. 3(e)). In an experimental sample, the differences in the PSFs between the four types of Qdots are also apparent (Figs. 4(a) and 4(b), & Visualization 3 for 2-color discrimination with the optimized PSF). The net is able to readily classify fields of view with high efficiency (Fig. 4(c)). Tested over patches extracted from five FOVs, the net correctly predicted the color of $96.8\% \pm 2.1\%$ Qdots ($N_{\text{total}} = 365$, where the percentage is the weighted mean and the uncertainty is represented by the weighted standard deviation calculated from the 5 FOVs), improving over the $81.4\% \pm 5.3\%$ for the normal PSFs recorded in the same ROIs (Fig. 4(d)). The mean signal levels for the four colors were: 3130, 8150, 8180, & 2132 photons for the 565, 605, 705, & 800 QDs, respectively, and the background standard deviation of the time-averaged images used for classification was 3.5 photons per pixel.

While immobile emitters can be readily distinguished by the net, the most useful application for simultaneous multicolor imaging is in time-evolving samples, *e.g.* particle-tracking experiments.

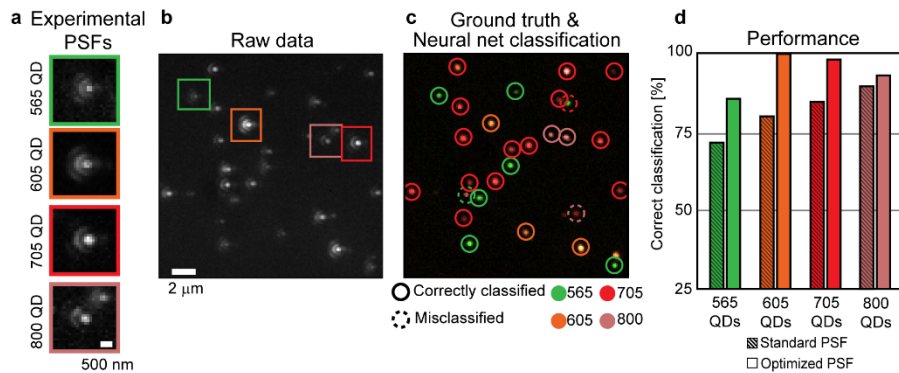


Fig. 4. Four-color classification of emitters using the optimized phase mask. (a) Experimental images of four types of Qdots from a larger FOV (b). (c) Classification of Qdots in the same field of view (circles) overlaying an image of emitters which appeared in the raw data, artificially colored according to their ground-truth wavelengths. Closely-spaced Qdots, such as in the lower right, were compared to the brighter of the two emitters in the GT image. (d) Performance of color determination for the normal and optimized PSFs ($N = 60, 120, 156, 29$, respectively).

Unlike static samples, tracking experiments have the added challenge caused by motion blur: the PSF in each frame is convolved with the trajectory within the exposure time. In brief, samples for tracking experiments were prepared by squeezing a 1 μL droplet containing two types of fluorescent, sub-diffraction-sized beads between a coverslip and glass slide (red beads: 645/665 nm absorption/emission; and green beads: 505/515 nm absorption/emission). The sample was then imaged with a 20X Air objective, NA 0.75, producing a quasi-2D diffusion chamber $\sim 4 \mu\text{m}$ in height, which was similar to the depth-of-field of the imaging system (Fig. 5(a)).

To gather training data for the net, three types of movies were recorded sequentially with a constant and net-determined optimized SLM pattern: images containing both colors of beads (Visualization 4, Appendix 5.6.), as well as red-filtered and green-filtered images, similar to the Qdot experiments. Unlike the immobile Qdot data, the diffusing beads yielded a more time-varying data set that can be treated as many independent measurements of the PSF. The nets were trained using sequential frames of re-centered ROIs around each bead (Fig. 5(b)) obtained in 40 separate movies, totaling 577 red and 973 green beads for the standard PSF and 299 red and 512 green beads for the optimized PSF. In each frame, the mean signal was 17,420 and 18,580 photons for red and green beads, respectively, with a background standard deviation of 37 photons per pixel. Due to the time-varying PSF, multiple frames were simultaneously fed to the net to determine the color. To evaluate the performance, the nets were used to determine the color of 159 beads imaged in 9 movies with the standard and optimized PSF (Fig. 5(c)). In all conditions tested, the optimized PSF outperformed the standard PSF, *e.g.* using 20 frames to determine the color, the standard/optimized-PSF nets correctly classified $84.3 \pm 6.0\%$ and $93.1 \pm 4.4\%$, respectively, and the percentages and uncertainties were computed as described previously.

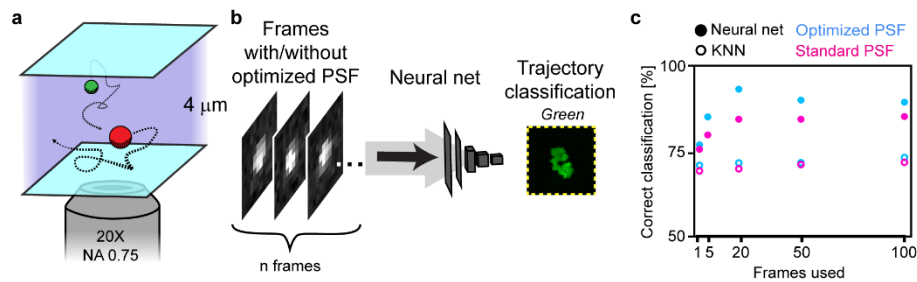


Fig. 5. Color determination of moving microspheres. (a) Illustration of the diffusion chamber. (b) Schematic of the neural net classifying sequential groups of frames belonging to the same emitter as red or green. (c) The performance of the net as a function of the number of frames used for classification.

Not all beads in the sample were mobile, and the net predicted the identity of static and mobile particles with different, albeit with high efficiencies. Using the standard PSF, 88.7% of the 97 mobile and 76.2% of the 62 static beads were correctly predicted. Using the optimized PSF, 92.8% of the mobile beads and 93.6% of the static particles were correctly predicted.

The neural net significantly outperformed nearest-neighbor analysis for both the standard PSF, and an optimized PSF (open circles, Fig. 5(c)). While a single image alone was not typically sufficient to determine the color, inputting multiple frames for the net enabled more reliable color prediction. As expected, the intentionally-designed, wavelength-dependent PSF enables color prediction from fewer frames.

3. Discussion

Neural networks have been shown to constitute a powerful tool in microscopy posed to replace existing algorithmic approaches [25,34,35]. Here, we have demonstrated not only how deep learning is capable of alleviating the need for physical spectral-filtering components, but importantly can be used to design the optical system itself. This approach resolves a key challenge in PSF engineering, how to best formulate the optimization cost function.

Localization microscopy is particularly well suited for deep learning because emitters form a relatively homogeneous population. This reduces the requirement for enormous sets of diverse training data which is often required for machine learning. Historically, localization microscopy relies on fitting only a few parameters to an image of an emitter (*e.g.* amplitude, Gaussian widths, astigmatism angle, and background, or for maximum-likelihood-estimation – all of the pixels in the region of interest of the PSF [36]), whereas neural nets rely on orders-of-magnitude more tuned parameters. This flexibility of the network architecture allows it to capture more subtleties in the data. For example, in the Qdot experiment using the standard PSF, the net far outperformed the alternative classification approaches (Fig. 2(h)).

A key feature in the standard PSF is the size-increasing blur incurred by defocus. Since the size of the diffraction limit (which determines the size of the PSF) scales linearly with wavelength, a size-only metric used to determine the identity could easily confuse defocused green emitters with in-focus red emitters. The *z*-range for correct identification using the standard PSF in an aberration-free, diffraction-limited system is therefore limited to the approximate depth such that the defocused green PSF will be smaller than the in-focus red PSF: ~several hundred nanometers (see Appendix 5.5.1.). In an experimental sample, the problem of defocus is further complicated by the wavelength-dependent focal length of even chromatically corrected objectives which can make the PSFs appear more similar in size (Appendix 5.8.), also observed by Cabriel *et al.* [28]. Unlike a comparison of the Gaussian-fit size, the net utilizes any differences in the PSFs, and therefore any chromatic aberrations in the optical system can help break the ambiguity between the two shapes. In the experiments

described here, this subtlety was well captured by the net in the Qdot data; however, utilizing the small differences in PSFs is a bigger challenge in low SNR samples, such as in single-molecule localization microscopy in cells. Here, the 3D nature of biological samples makes the single-channel, standard-PSF net determination only applicable to areas of the sample that are thin, and the noise may mask the PSF subtleties used for higher-signal experiments, such as the periphery of cultured cells (See Appendix 5.7.).

To further improve the differentiation between emitter colors, we employed PSF engineering to explicitly encode wavelength-dependent differences into the shape of the image, making it more robust to defocus and other confounding factors (Appendix 5.5.). Importantly, by training the neural net to identify the emitter colors while simultaneously training the SLM-optimizer to find the best PSF-shaping phase mask, the problem of determining the optimal PSF can be simplified to converge on the pattern that optimizes for the result we want: precise localization and color differentiation. Interestingly, for various numbers of colors, the net converged on qualitatively similar phase masks: a smooth mask with a discontinuity (Appendix 5.2.6.).

To compare the output of the net to other solutions (Appendix 5.4.2.), we examined the performance versus several other reasonable PSFs. First, for each PSF, we trained a Reconstruction net to analyze images of simulated emitters at the focal plane and randomly distributed over a 500 nm z-range, and assessed the performance in terms of emitter detection, localization precision, and color differentiation. While the alternative PSFs improved over the standard PSF in these metrics, the net-optimized mask showed more robustness while maintaining high detectability. Interestingly, the optimized PSF showed no degradation of localization precision relative to the standard PSF for the signal & background regime obtained for the Qdot data; however, at low SNRs, the SLM-optimizer converges on a PSF that is similar to the standard one (Appendix 5.5.6.).

To test our method in the regime useful for super-resolution localization microscopy, we evaluated simulated images at different SNRs and densities using the optimized PSF (Appendix 5.5.2. and 5.5.3.). Single-molecule, super-resolution data sets consist of many frames containing unique sets of emitters. To reconstruct a super-resolved image, as many molecules as possible must be localized. Thus, for a given acquisition time, increasing the density of emitters in each image directly translates into a better reconstruction, so long as the emitters can still be localized with high precision [25,37–41]. We simulated overlapping emitters and found that the Reconstruction net performed well up to ~ 8.9 emitters/ μm^2 , which is a medium-to-high density of emitters. To assess the performance with a low SNR, we simulated emitters on a noisy background and measured the localization and color-determination efficiency. With 200 mean background photons per pixel, we found that the net could identify emitter colors even in adverse conditions, but performed best when signal photons exceeded 800.

Data sets consisting of static emitters with low SNR and high density are challenging for the net, but the PSF shapes for each class of emitter are similar. Images containing randomly moving particles, however, can vary significantly due to the stochastic step size in diffusion and associated motion blur that is convolved with the PSF. Surprisingly, the neural net is able to distinguish between two types of fluorescent beads from relatively short trajectories. When using an optimized PSF, the net achieves further improved performance, demonstrating the applicability of our approach for time-evolving samples (Fig. 5(c)).

To optimize discrimination between emitters, we have shown that PSF engineering can be done in coordination with net training to maximize on the strengths of the reconstruction net. Here we have focused on color discrimination, however, PSF engineering could also be used for analyzing any or multiple effectors on the shape of the PSF, *e.g.* z-position, molecular orientation, movement dynamics, and number of contributing emitters.

4. Methods

4.1. Deep learning

Localization of emitters was performed either using the ImageJ [42] plugin ThunderSTORM [43] or a custom peak-finding algorithm implemented in MATLAB (Mathworks).

The deep neural nets were implemented in MATLAB and several functions of the MatConvNet package [44]. Our setting required two types of nets: one for color determination, and one for simultaneous PSF optimization and color determination with the resulting PSF. Color determination is in fact a conventional classification task. For our color-determination net, we adopted an architecture similar to that proposed by Ledig et al. [45]. Our net contains 9 convolutional layers with an increasing number of channels and a decreasing spatial resolution (implemented by three stride 2 convolutions), followed by two dense (fully-connected) layers. The last layer of the net is a sigmoid layer that operates over a scalar for the two-color determination net and a sigmoid layer that operates over a vector of length four for the four-color determination net. The outputs represent the two/four color probabilities. We used the cross-entropy loss and Adam optimization for training this net [46].

Determining an optimal microscopic PSF through the design of an SLM phase pattern is a unique task that has not been addressed in previous work. Here, our architecture comprised an SLM-optimizer which generates images based on a randomly-placed simulated emitters at given wavelengths in addition to a phase-delay pattern, and a reconstruction net which classifies emitters based on the PSFs induced by the learned SLM. Both components were trained simultaneously to learn to optimize voltages of each of the SLM's pixels and subsequently perform color determination. In neural net terminology, the SLM's voltage mask can be thought of as a nonlinear layer, whose parameters have to be optimized, similar to conventional linear layers.

The SLM-optimizer selects the voltage value for each pixel among 50 possible voltages, similarly to the principle used by Chakrabarti [22]. The induced PSFs are then convolved with simulated point sources and the resulting grayscale image is subsampled and degraded by Poisson noise.

The color reconstruction net contains eight convolutional layers, followed by two deconvolution layers, each doubling the spatial resolution (Appendix 5.1.). It also contains a skip-connection branch with one deconvolution layer that increases the spatial resolution by 4 and bypasses nine of the convolutional layers of the main branch. This is done to alleviate the gradient-vanishing problem [47]. The outputs of both branches enter a last convolutional layer followed by a sigmoid layer. The net's output is a probability map for each pixel of the gray image being one of the four colors on a high-resolution grid. The loss is again cross-entropy. Additional details on the net architecture and training are in Appendix 5.1.

4.2. Microscopy

4.2.1. Imaging system

Imaging experiments were performed using a standard inverted microscope (TI Eclipse, Nikon), equipped with an XY Proscan III translational stage and a Nano-Z Piezo stage (both Prior Scientific) and sCMOS camera (95B Prime, Photometrics). The instrument was controlled with Nikon Imaging Software and illuminated with a fiber-coupled laser-light source (iChrome MLE, Toptica). To allow for the placement of additional optics (*i.e.* the spatial light modulator, PLUTO-VIS, Holoeye, after a linear polarizer, Thorlabs), the imaging path was extended with two, $f = 15$ cm lenses (Thorlabs, Fig. 6).

When the SLM is inactive (*i.e.* turned off), the SLM acts as a mirror, recapitulating the standard PSF behavior of a conventional microscope configuration (Fig. 2(a)). When the SLM is turned on, the PSF is modulated. Experimental results using the standard PSF were

verified in the normal microscope configuration, but for quantitative comparison, the extended system was used for all measurements presented.

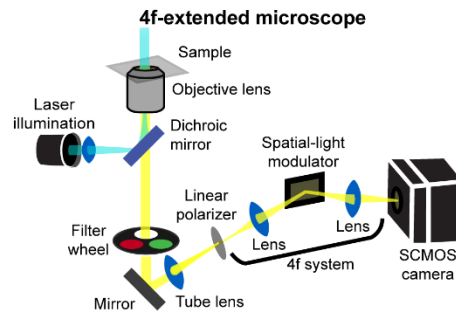


Fig. 6. Microscope schematic using an extended imaging path.

4.2.2. Immobile quantum dots

For two-color Qdot experiments (Fig. 2), 565 and 705 nm emission-peak nanoparticles (Life Technology) were diluted in 1% PVA and spin coated onto a 1.5 coverslip slide (Thermofisher) achieving a final density of $0.08/\mu\text{m}^2$ green Qdots, 0.05 red Qdots/ μm^2 . Samples were then excited with 405 nm light and imaged through a Nikon 100X NA 1.49 TIRF objective in Epi-illumination mode. The emission light was chromatically filtered with a dichroic and long-pass filter (ZT488rdc & ET500LP, Chroma) to remove background and scattered illumination light. To distinguish the true color of the Qdots recorded with the grayscale camera, an additional 565/70 or 650 LP filter (both Semrock) was inserted in the imaging path. For the matched-filter comparison, the mean PSF was generated by subpixel-shifting each ROI containing a Qdot according to the localized position in a field of view, normalizing each image and taking the average. In the data set, each PSF was then shifted according to its localized position and the correlation with the mean red and green PSFs were compared. In the four-color experiments (Fig. 4), the sample was prepared with 565, 605, 705 and 800 nm Qdots. To obtain the multicolor image, a dichroic mirror was used in combination with a notch filter and long-pass filter (ZT488rdc, ZET405NF & ET500LP, Chroma) and a 2.5 minute acquisition with 50 ms frames was recorded with and without the SLM activated. To attain ground truth images, four bandpass filters were used sequentially for 30 second exposures to identify the individual species, which were then removed for the final image (AT565/30m, AT625/30m, AT705/30m, AT800/40m, all Chroma). Due to blinking, emitters were only analyzed if they appeared in all three data sets (SLM active/inactive, and in one of the four spectrally filtered images).

4.2.3. Diffusing fluorospheres

For the diffusing-bead experiment (Fig. 5), 100 nm green and 200 nm red fluorescently labeled microspheres (Life Technology) were diluted into 40% glycerol in water (v/v). From the mixture 1 μL was then pipetted onto a glass coverslip and pressed onto a glass slide and sealed with clear nail polish. Both surfaces were pretreated with a ~ 20 mg/mL casein solution to decrease the propensity for sticking of the fluorescent beads to the glass. In most regions of the sample, the beads remained in solution for several hours; however, there were areas of the sample where the passivation layer was flawed and the majority of beads had adhered to the surface. Three fluorescent filters were used in combination with different excitation-laser combinations to image the green, red, and both beads at once. Green-bead images were recorded with a 488 nm excitation and a green filter set (ZT488rdc & ET500lp, EM525/50bp, Chroma); red beads were imaged with a 650 nm laser and a red filter set (ZT650rdc, EM650lp); images of both beads were done using a multi-bandpass filter set (ZT405/588/561/647rpc, ZET405/488/561/647m, Chroma). All imaging was done using a

Nikon 20X air objective, NA 0.75 without the additional 4f extension used in the Qdot experiments.

5. Appendix

5.1. Architecture and hyper parameters of the color determination nets

The architecture and training procedures for the two-color determination nets were all similar, *i.e.* two-color Qdot differentiation, diffusing emitters, and single-molecule blinking experiments (Fig. 7(a)). The four-color classification nets used a slightly modified architecture (Fig. 7(b)).

The first step in training the net is to attain a suitable image library. For the two-color nets, patches containing (11×11) pixels² centered on emitters are extracted from the larger image data. A randomly selected fraction of these patches is used in each training iteration unchanged; the remaining patches go through an augmentation process to ensure that the net remains broadly applicable to varying SNR conditions.

For the two-color, immobile Qdot net, the training data was split into two even parts. For the augmented half, the median background gray level of each patch was subtracted and a random background floor level in the range [400, 800 photons] was added together with a Poisson noise with a random variance, λ , in the range of [16, 64 photons].

For the moving-emitter net, 20 out of 1000 centered 15×15 patches were randomly selected for the augmentation. The median background gray level for each patch was subtracted and a random background floor level in the range [100, 600 photons] was added together with a Poisson noise with a randomly selected λ in the range of [225, 784 photons].

Patches were randomly shifted in the horizontal and vertical directions in the range of [-3, 3 pixels] and enlarged to a 16×16 patch by replicating the borders of the smaller patch.

The detailed architecture of the two-color determination net shown in Fig. 7(a) is described in detail below. The first Batch-Normalization (BN) layer in the net is used to normalize the data. Its gains and biases are set to ones and zeros and are not learned parameters.

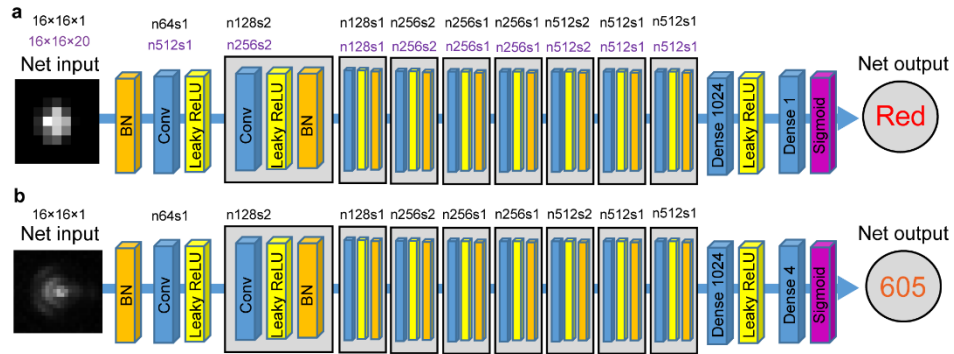


Fig. 7. Architecture of the color determination nets. (a) Two-color determination net, where the number of feature maps (n) and stride (s) of each convolutional layer are written for each experiment. Black denotes the Qdot net for the standard PSF and the super-resolution net. Purple text describes the net for moving beads. (b) The four-color determination net.

All the convolutional weights of the net are 3×3 in size.

A dropout with $p=0.5$ is implemented after each Leaky ReLU layer ($\alpha=0.01$), except for the last one, and L_2 regularization is used with $\lambda=10^{-6}$.

The Adam optimizer is used with $\beta_1=0.9$, $\beta_2=0.999$, and $\varepsilon=10^{-8}$ to update the net's parameters. We use a cross-entropy Loss defined by Eq. (1), where n is the batch image index, N is the batch size, GT is the ground truth (0/1 for a red/green emitter) and Z is the net's output.

$$LOSS = \frac{-1}{10 \cdot N} \sum_n [GT \cdot \log(Z + \varepsilon) + (1 - GT) \cdot \log(1 - Z + \varepsilon)] \quad (1)$$

The nets were trained for various numbers of iterations at decreasing learning rates (Table 1).

Table 1. Training parameters for the two-color nets

	Qdots	Mobile-emitters	Single-molecule
Iterations – Learning rate	28K – 0.0001 476K – 0.00001	100K – 0.0001	158K – 0.0001
Batch size	32	32	16

The architecture and training procedures for the four-color determination nets were similar for the standard PSF and the optimized PSF (Fig. 7(b)). In both, 15×15 pixel² patches were extracted from larger images. For the augmented half of both the standard PSF and optimized PSF datasets, the median background gray level of each patch was subtracted and a random background floor level in the range [100, 190 photons] was added together with a Poisson noise with a random variance, λ , in the range of [0.5, 10.5 photons].

The patches were then randomly shifted in the horizontal and vertical directions in the range of [-4, 4] pixels and enlarged to 16×16 pixel² patches by replicating the borders of the original extracted patches.

All other net parameters and Loss function were identical to those of the two-color net, although the ground truth (*GT*) is now a vector of size 4, according to the 4 possible colors.

Similar to the two-color nets, the four-color nets were trained in multiple stages with varying iteration numbers and learning rates (Table 2).

Table 2. Training parameters for the four-color nets

	Standard PSF	Optimized PSF
Iterations – Learning rate	38K – 0.001 60K – 0.0005	20K – 0.001 100K – 0.0005
Batch size	32	32

When training the net, it is important to consider the possibility that the training data may contain some bias that is conferred to the net *i.e.*, if more patches of one type of emitter are used versus the other, the net may be more likely to predict that color. Using a suitably large training dataset as well as our augmentation process, which further de-homogenizes the data is used to help alleviate such confounding factors.

5.2. Architecture and hyper parameters of the SLM-optimizer and Reconstruction net

To find the best voltage pattern for an SLM for neural-net localization & classification, we trained two systems in tandem, an SLM-optimizer and a Reconstruction net. The architecture for the SLM-optimizer and associated Reconstruction net are shown in Fig. 8.

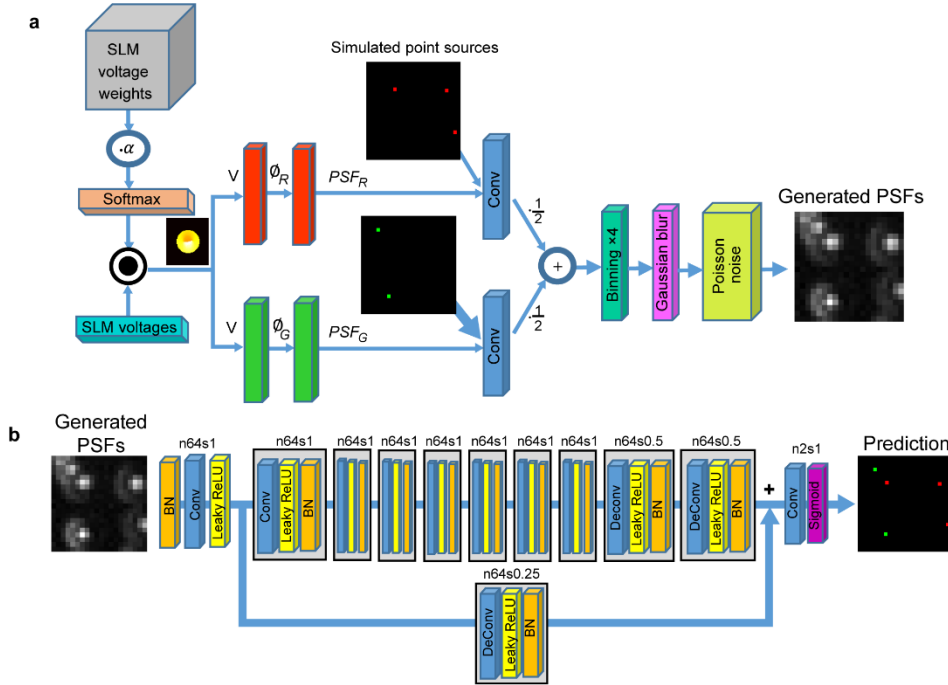


Fig. 8. Optimizing voltage patterns on an SLM using neural nets. (a) The network architecture for SLM-optimizer. (b) The Reconstruction network architecture.

5.2.1. SLM-optimizer

The SLM-optimizer's purpose is to simulate libraries of experimentally-acquired images of emitters at random positions within a field of view after encountering an SLM with a particular phase pattern (see Figs. 3(b) and 8(a)). The SLM-voltage weights are the only learned parameters. There are $215 \times 215 \times 50$ parameters, where 215×215 represents simulated area of the back focal plane, and the depth represents all the possible SLM voltages. Note that we only use a subsection [12, 61 DU] out of the full possible SLM settable values of [0, 255 DU] for improved correspondence between simulation and experiment. These weights are then multiplied by a scalar parameter, α , which is slowly increased as iteration number increases, according to Eq. (2), where t is the current iteration number and $\gamma = 2.5 \times 10^{-5}$.

$$\alpha(t) = 1 + \gamma \cdot t^2 \quad (2)$$

The result is passed through a Softmax layer, which normalizes it such that each one of the 215×215 SLM pixels contains a probability vector for each one of the possible 50 SLM voltages; note that pixels outside the area covered by the BFP do not modify the PSF and are therefore not adjusted by the net during optimization (Visualization 2). The role of $\alpha(t)$ is therefore to make the probabilities sharper as the iteration number increases. Next, an inner product is implemented between each one of the voltage-probability vectors and a vector of the corresponding voltages. This operation chooses one voltage value for each SLM pixel, and the result is a 215×215 SLM voltage pattern. The SLM pattern is then used for the red and green PSFs channels.

5.2.2. The green-PSF channel

The imparted phase for green radiation ($\lambda = 565$ nm), is obtained from the voltage pattern by using the SLM response curve (Fig. 3(a)). The green PSF is then produced as shown in Eq. (3):

$$PSF_G = \left| \mathcal{F}^{-1} \{ circle_G \cdot e^{i\phi_G} \} \right|^2 \quad (3)$$

Where PSF_G is the green PSF, \mathcal{F}^{-1} represents the 2D spatial inverse Fourier transform and $circle_G$ is a centered circle. This circle is present due to the ‘cone’ of light rays that are collected by the microscope from an emitter and projected onto its pupil plane and then on the Fourier space of the 4f system. The physical diameter of the circle in the SLM Fourier plane is approximated by Eq. (4):

$$D = 2 \cdot f \cdot \frac{NA}{M} \quad (4)$$

where D is the diameter of the back focal plane image, f is the focal length of the first lens in the 4f system, NA is the numerical aperture of the objective, and M is the image magnification. For simulations we use a high resolution grid with a $4X$ reduced pixel size relative to the camera’s pixel size. The Fourier space size in simulation can be obtained by Eq. (5):

$$A_G = \frac{\lambda_G \cdot f}{pixel_{HR}} \quad (5)$$

Using all the physical sizes of our optical system, $f = 150$ mm, $NA = 1.49$, $M = 100$, $\lambda_G = 565$ nm (mean wavelength for the green Qdots used in experiments), $pixel_{HR} = 2.75$ μ m, we can confirm that the circle diameter is 14.5% out of the green SLM space size $\left(\frac{D}{A_G} \right)$, or 31 pixels out of the 215 green SLM size that was chosen arbitrarily in the simulation.

5.2.3. The red-PSF channel

The red PSF is produced similarly to the green as Eq. (6):

$$PSF_R = \left| \mathcal{F}^{-1} \{ circle_R \cdot e^{i\text{Padding}(\phi_R)} \} \right|^2 \quad (6)$$

Using $\lambda_R = 705$ nm (mean wavelength for the red Qdots used in experiments), we can confirm that the circle diameter (which is independent on the wavelength) is 11.6% out of the red SLM space size $\left(\frac{D}{A_R} \right)$, which means that the red SLM space size in the simulation is

$$\frac{31}{0.116} = 267 \text{ pixels.}$$

The padding in (Eq. (6)) is a zero padding operator that resizes the 215×215 ϕ_R phases to 267×267 .

5.2.4. The recombined green and red channels

Next, the net produces a high resolution gray image described by Eq. (7), where $Sources_R$ & $Sources_G$ are the high resolution grid images of the red & green emitters' locations and '*' denotes convolution.

$$Gray_{HR} = \frac{PSF_R * Sources_R + PSF_G * Sources_G}{2} \quad (7)$$

We use in the simulations 30×30 detector's grid patches in which a random number between [5,10] red and green points are located in random positions over the 120×120 high-resolution grid. Each point is assigned with a random signal value between [6000, 12000 photons]. In principle, any high-resolution grid size could be chosen at the cost of computation time.

$Gray_{LR}$ is obtained by passing $Gray_{HR}$ through a binning operation (performed by a 4×4 mean pooling layer multiplied by 16) and a convolution with a Gaussian with a standard deviation of 0.5 pixel that simulates a mild optical blur in the optical system. The last stage is obtaining the images by adding a background and contaminating with Poisson noise (Eq. (8)), where the background term is a constant image with a random gray level in the range [144, 676 photons].

$$Image = Poissrnd\{\lambda_{Poisson}\} = Poissrnd\{Gray_{LR} + background\} \quad (8)$$

5.2.5. The Reconstruction net

The reconstruction net architecture is presented in Fig. 8(b). The input to this net is the low resolution 30×30 image of generated PSFs. Its purpose is to generate the high resolution 120×120 localizations and color determination maps.

The batch size is 16 and the Adam optimization is used with the same parameters as in the color determination net. No regularization is used.

We use a weighted cross-entropy loss defined by Eq. (9), where n is the batch image index, N is the batch size, and GT is the ground truth: a $120 \times 120 \times 2$ image for one batch image.

$$LOSS = \frac{-1}{10 \cdot N} \sum_{h,w,d,n} [Mask \cdot GT \cdot \log(Z + \epsilon) + Mask \cdot (1 - GT) \cdot \log(1 - Z + \epsilon)] \quad (9)$$

The first layer is a 120×120 grid of red predictions, consisting of 1 or 0 values that denote the existence of an emitter in each pixel, and the second image is the green predictions. Z is the net's output, and the sum is over all of its dimensions.

The *Mask* is a $120 \times 120 \times 2$ image consisting of scores. Its purpose is to encourage the net to correctly predict pixels that contain emitters by assigning them a higher score. Producing the *Mask* involves assigning each emitter in the red emitters' locations image a score of $\frac{3}{N_{red}}$, where N_{red} is the total number of red emitters, and each emitter in the green

emitters' locations image a score of $\frac{3}{N_{green}}$ where N_{green} is the total number of green emitters.

Then, the two images are arranged to produce the *Mask*. Finally, all the pixels in the *Mask* that are zeros are set to $\frac{1}{N_{zeros}}$, where N_{zeros} is the total number of zeros in the *Mask*. This

process provides the net a total score that is 3 times greater when it correctly predicts all the red and all the green emitters instead of correctly predicting all the pixels that don't contain emitters. Finally, the pixels in the *Mask* that are outside the central 80×80 central pixels are set to zero, so emitters near the borders do not contribute to the loss.

The optimized SLM estimation net for 2 colors was trained for 270K iterations with a learning rate of 0.0001 (Fig. 9).

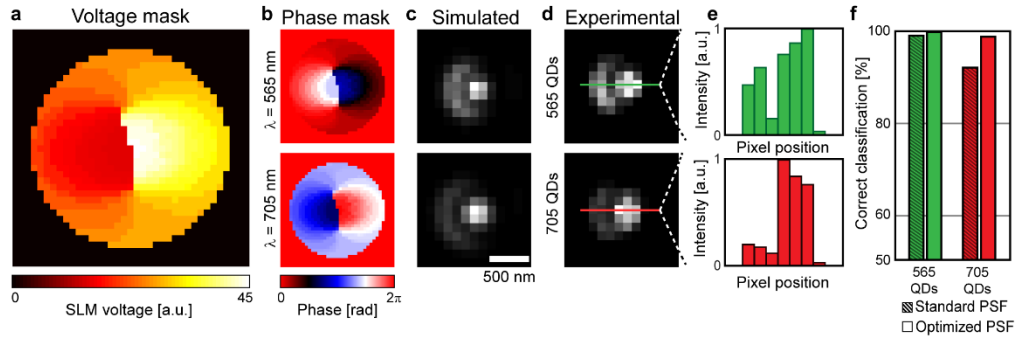


Fig. 9. Optimized 2-color PSF. (a) The optimized SLM-voltage pattern for color determination by a neural net. (b) The phase delay imparted to 565 and 705 nm light. (c) Simulated PSFs. (d) The experimental PSFs measured with Qdots. (e) Pixel values of the cross section from experimentally measured PSFs. (f) Performance of color determination for the normal and optimized PSFs.

In the two-color problem, the net converges to an SLM-voltage pattern shown in Fig. 9(a), where the two distinctive phase-delay patterns depend on the wavelength (Fig. 9(b)). Experimental PSFs, recorded with Qdots, match the simulations (Figs. 9(c) and 9(d)). The difference in the normalized intensity distributions are observable in the image cross-sections (Fig. 9(e)). To compare the reconstruction net's performance with the optimized PSF to the standard PSF, we repeated the experiment described in Fig. 2. For each FOV, an additional image was taken with the optimized SLM pattern (when the SLM is inactive, no PSF-modification occurs). The net correctly predicted the color of $99.4\% \pm 0.5\%$ Qdots ($N_{\text{total}} = 2195$, where the percentage is the weighted mean, and the uncertainty is represented by the weighted standard deviation estimated from six FOVs), improving over the 96.4% for the normal PSF (Fig. 9(f)).

5.2.6. Optimized SLM patterns

Interestingly, the 5-color problem also produced a similar pattern as the 2 and 4 color nets (Fig. 10).

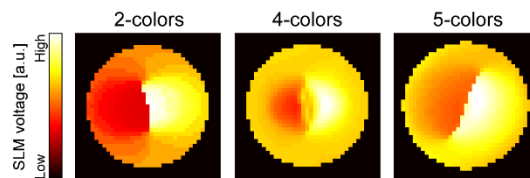


Fig. 10. Optimized voltage masks for 2, 4 & 5-color classification.

5.3. Separability of emitters

5.3.1. t-distributed stochastic neighbor embedded projections

t-SNE is a useful tool to visualize the separability of data sets into clusters [29]. To determine whether asymmetric 2D-Gaussian fits were indeed separable by some measure, we performed t-SNE analysis on the outputs of our localized values (Fig. 11(a)). While there was no clear separation between red and green Qdot data from the localization data, the analyzed 11×11 pixel² ROIs did show clustering *via* t-SNE analysis over a broad range of parameters for both the standard and optimized PSFs (Figs. 11(b) and 11(c)).

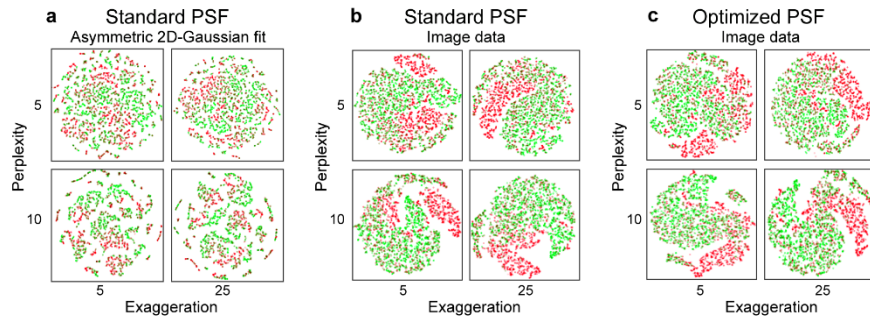


Fig. 11. t-SNE projections showing separability of 2-color Qdot data. Red and green spots represent the colors of the Qdots. (a) Analysis of the asymmetric 2D-Gaussian fit parameters shown in Fig. 2(g). (b) t-SNE projection of 11×11 pixel² patches containing red and green Qdot emitters and (c) 11×11 pixel² using the optimized PSFs. For both b & c, ten principle components were used; however, a wide range of selected parameter-values similarly showed separability of the two types of Qdots.

5.3.2. Testing other classifiers

In addition to the neural network, we tried classification tools such as a global threshold over emitter's brightness, K-nearest neighbors (KNN) applied on the 2D astigmatic Gaussian parameters (σ_1 , σ_2 , θ) with 81 neighbors and finally, KNN with 21 neighbors applied on the 11×11 pixel² patches themselves (Table 3). In each case, the number of neighbors, K , was chosen to maximize classification success. The data is also shown in Fig. 2(h).

Table 3. Results of different classifiers

Classifier Success	Results [%]
Neural net	96.4
KNN over patches	76.7
Threshold over brightness	70.3
KNN over Gaussian parameters (σ_1 , σ_2 , θ)	67.3
Matched Filter	55.0

5.4. Comparison to other PSFs

The utility of a multicolor, localization-microscopy method hinges on the ability to differentiate colors while maintaining the localizability of emitters. In this section, we first compare the localizability of the standard PSF and the net-optimized PSF on augmented experimental data. Next, we compare the net-optimized PSF to other possible solutions by simulation.

5.4.1. Experimental assessment of localization precision and classification for the optimized and standard PSFs

To assess the localization precision with the standard and net-optimized PSFs, several ROIs from the four-color classification experiment described in Fig. 4 were selected. Each ROI contained a Qdot that fluoresced continuously for nearly the entire duration of the experiment (with and without the SLM activated). Frames where the Qdot blinked off were removed from subsequent analysis. Each ROI was then averaged for 250 frames to generate a matched filter used to localize the position of the emitter in the individual frames. The data set was then augmented with a Poisson-distributed random variable [0,350 photons] to produce various SNR conditions (Fig. 12). The average signal level per frame fluctuated around [6200 photons] for the Qdot shown; however, the general trend is representative of all data analyzed. At low background, the localization precision for both PSFs were similar (in

agreement with simulations described in Appendix 5.4.2.). With worsening SNR, the localization precision of the net-optimized PSF degrades more quickly relative to the standard PSF, as to be expected when the signal is spread out over a larger PSF.

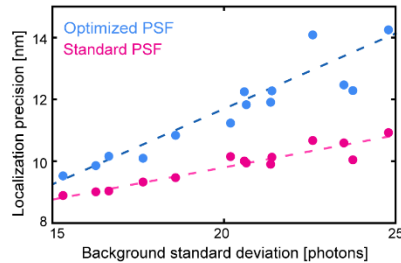


Fig. 12. Localization precision in various noise conditions for the standard and net-optimized PSFs.

5.4.2. Simulated comparisons between PSFs

To compare the net to other reasonable solutions for multicolor classification, we compared the 2-color optimized PSFs to three other PSFs: the standard PSF, a half-optimized PSF obtained after 900 iterations of the 2-color's net training, and finally, a 2-color astigmatic PSF. The latter was obtained by generating the directional astigmatic PSFs (using a combination of Zernike vertical astigmatism polynomial and the defocus polynomial at $Z = 140\text{nm}$ and $Z = -175\text{nm}$), and then obtaining a phase for each one of the colors by the Gerchberg-Saxton phase retrieval algorithm. Finally, a voltage mask is obtained by solving a LS problem [18]. The four 2-colors PSFs are presented in Fig. 13. We also compared performance of the 4-color optimized PSFs (Fig. 3(e)) and standard PSFs.

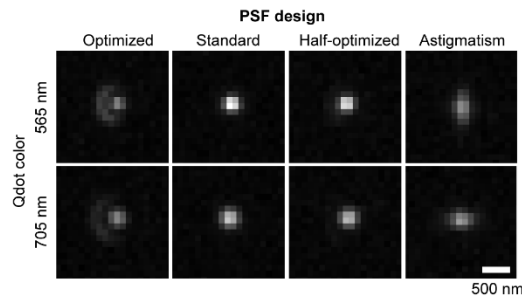


Fig. 13. Two-color PSFs used for quantitative comparison.

To compare the performance between PSFs, we defined several evaluation metrics: *detection*, *color-determination*, *false alarms*, and *localization precision*:

The detection metric: percentage of detected emitters within 69 nm of their true position, even if the color determination was wrong.

The detection and color-determination metric: percentage of detected emitters within 69 nm of their true position whose subsequent color-determination is also correct.

False alarms: average number of emitter detections incurred per pixel, calculated from pixels containing no emitters.

Localization precision: averaged distance between a detected emitter and the true position, calculated over all detected emitters.

For the 2-color simulation [5–10], emitters of each color were generated with [5–10K] signal photons and an added background [144, 676]. For the 4-color simulation, [30–60K] signal photons were used, and FOVs contained 1–2 emitters of each color (4–8 total) with randomly added background in the range [144,676].

The results are presented in Table 4. The optimized PSFs achieved the highest performance. The localization precisions for the standard and optimized PSFs were similar at this SNR; however, in cases where the emitters overlapped, the optimized PSF shows superior performance because the net is better able to discriminate and localize the two emitters simultaneously.

Table 4. Comparison of PSFs in focus

	2-color PSFs				4-color PSFs	
	Optimized	Standard	Half-optimized	Astigmatism	Optimized	Standard
Detection [%]	99.9	99.9	99.8	97.9	100	100
Detection & Discrimination [%]	99.7	98.1	98.2	94.3	99.6	97.9
False Alarms [%]	0.02	0.04	0.04	0.09	0.05	0.04
Localization [nm]	4.9	5.5	5.5	16.5	5.5	5.5

5.5. Simulation-based evaluation of the net under various conditions

5.5.1. Testing z-range

To assess the performance under varying focus- position, a graph of the PSF width vs. focal position was obtained for the 4-color standard PSFs (Fig. 14). It can be seen that up to ~250 nm change in the focal position, the 4 colors maintain their size order and are thus distinguishable.

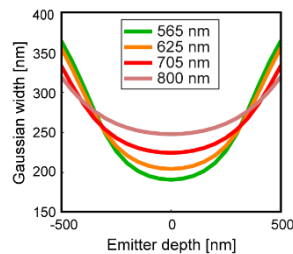


Fig. 14. Simulated size of the PSF with defocus.

We repeated the test described earlier for simulated emitters, but allowed emitters to appear over a 500 nm z-range for the same 2 and 4-color PSFs. The same nets used in the planar case were used to classify the emitter colors without any retraining. In both the 2 and 4-color cases, the optimized masks performed best (Table 5). Note that as the z-positions increase, the net classification of the astigmatic PSF becomes problematic since the encoded defocus used for classification will become ambiguous with the z-position.

Table 5. Comparison of PSFs over a 500 nm z-range

	2-color PSFs				4-color PSFs	
	Optimized	Standard	Half-optimized	Astigmatism	Optimized	Standard
Detection [%]	99.8	99.8	99.8	97.6	99.8	99.8
Detection & Discrimination [%]	99.7	97.5	97.5	81.7	99.7	97.5
False Alarms [%]	0.03	0.05	0.05	0.12	0.03	0.05
Localization [nm]	6.9	7.7	8.0	19.5	6.9	7.7

5.5.2. Testing emitter densities

To assess the net's performance in a variety of experimental-like conditions, images were simulated with various emitter densities and SNRs (Fig. 15). For these simulations, we fix the optimized SLM pattern in the optimized SLM estimation net generated for the two-color sample and only update the reconstruction net's weights. The reconstruction net is trained with a random number of red and green emitters between [2,20] so that the total number of emitters is between [4,40] in each $2.2 \times 2.2 \mu\text{m}^2$ FOV (20×20 pixels). We also assign random signal values for red and green emitters in the range [12000, 24000 photons]. All other net's parameters are the same as in section 2. We initialize the reconstruction net weights to the weights learnt in section 2 and train for 100K iterations with a training rate of 0.00005.

For testing, at each emitter density, 100 simulated images are generated with a random number of red and green emitters (the total number of emitters corresponds with that density). Next, the reconstruction net is used in order to generate localization and color determination maps. Finally, we perform a post processing which leaves in those maps only the pixels that are local maxima, threshold the result at 0.985, calculate the Detection, Detection & Determination, False Alarms and Localization precision measures and average them over the 100 examples.

An example region with 8.9 [emitters/ μm^2] density (43 emitters in a $2.2 \times 2.2 \mu\text{m}^2$ sample region, which correspond to a 20×20 pixel² ROI area on the detector) is shown in Fig. 15(b).

In a real experiment, the imaging system contains aberrations that may make a net trained on simulated data unsuitable (*e.g.* chromatic-dependent focus, Appendix 5.8.), necessitating the use of experimentally acquired images for training. Problematically, the ground truth is unobtainable at high emitter densities, where traditional methods fail. It is possible, however, to attain a high SNR data sets of sparse, bright emitters over many frames that can be localized and classified, then combine the images and augment an appropriate level of background noise as was shown previously [25].

5.5.3. Testing emitter SNRs

For each signal level, 100 images were generated with a Poisson-distributed background of mean 200 photons. One red emitter and one green emitter are placed in each image. The reconstruction net is used in order to determine the localization and color of identified emitters in the FOV. An example is shown in Fig. 15(d).

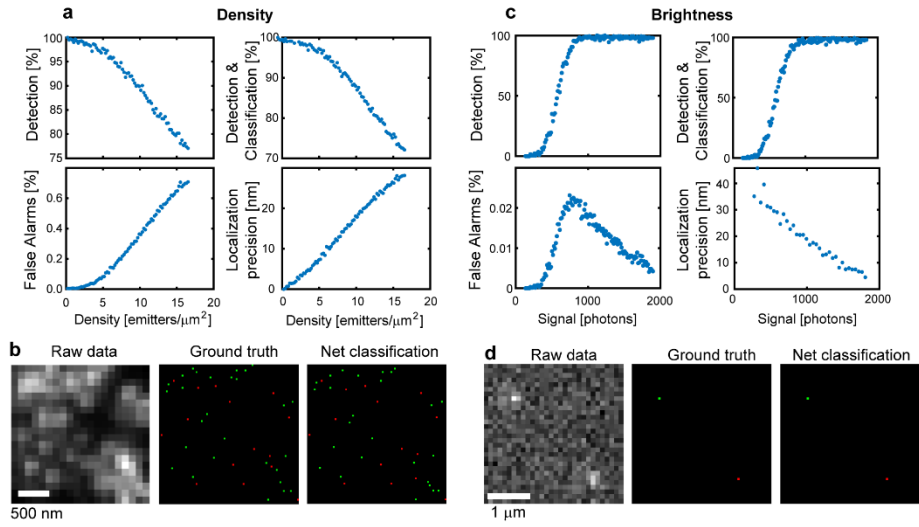


Fig. 15. Evaluating performance in various conditions. (a) Performance over different emitter densities (b) An example of a generated image and analysis with a density of 8.9 [emitters/μm²]. (c) Performance with various emitter brightnesses. (d) An example image and analysis of 800 signal photons per emitter and 200 background photons per pixel.

5.5.4. Evaluating net's performance over emitter-wavelengths proximity

Using simulations, we determined the net's ability to distinguish emitters of various wavelength differences (Fig. 16). First, we trained three different nets with similar parameters to those we used so far, for three wavelength proximities:

$$\Delta\lambda_A = 5 \text{ nm } (\lambda_1 = 565 \text{ nm}, \lambda_2 = 570 \text{ nm})$$

$$\Delta\lambda_B = 40 \text{ nm } (\lambda_1 = 565 \text{ nm}, \lambda_2 = 605 \text{ nm})$$

$$\Delta\lambda_C = 140 \text{ nm } (\lambda_1 = 565 \text{ nm}, \lambda_2 = 705 \text{ nm})$$

We also trained the reconstruction net with a constant SLM (no pattern) for emitters with the same proximities. For each simulation, we evaluated the average over 16K randomly generated images using the protocol described previously.

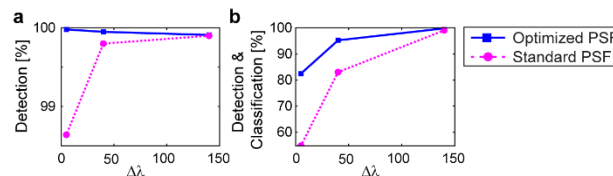


Fig. 16. Evaluating performance over emitters' wavelengths proximity. (a) Detection performance. (b) Detection & determination performance.

5.5.5. Evaluating net's performance over different number of emitters' colors

To determine the number of spectrally distinct emitter types, N_{types} , that could be simultaneously identified using an optimized SLM pattern (Fig. 17), we applied the following modifications to the SLM-optimizer described earlier: The SLM voltage is split into N_{types} color channels, each one of them generates one color PSF. The image is then obtained by Eq. (10):

$$gray_{HR} = \frac{1}{N_{types}} \cdot \sum_{n=1}^{N_{types}} PSF_i * Sources_i \quad (10)$$

In the reconstruction net, the weight of the last convolutional layer consists of N_{types} filters, which means that the net's output has the size of $120 \times 120 \times N_{types}$, where GT and $Mask$ are of the same size.

We used $N_{types} = 2, 4, 5$. One or two emitters per color are simulated per image. The 2 color net was described previously. The 4 and 5 channels nets are assigned a random signal value between [30K, 60K photons] and the colors are [545 nm, 585 nm, 625 nm, 705 nm] for the 4 channels net and [545 nm, 585 nm, 625 nm, 655 nm, 705 nm] for the 5 channels net, corresponding to commercially available Qdots (Invitrogen). All the other parameters are the same as described previously. The 4-channel net was trained for 50K iterations with a learning rate of 0.0001 and a batch size of 8, and for 40K more iterations with a learning rate of 0.00005 and a batch size of 16. The 5-channel net was trained for 65K iterations with a learning rate of 0.0001 and a batch size of 8, and for 25K more iterations with a learning rate of 0.00005 and a batch size of 16. The performance was evaluated with 16K randomly generated images. In the 2 channel case, emitters were assigned a random signal between [18K, 36K photons].

The result of the four color determination experiment using the optimized PSF is presented in Fig. 4. The optimized SLM pattern was generated by simulation with the next parameters:

$f = 150$ mm, $NA = 1.49$, $M = 100$, $pixel_{HR} = 2.75$ μ m.

$\lambda_1 = 565$ nm: circle diameter is 14.5% out of the SLM space size, or 31 pixels out of the 215 pixels SLM size that was chosen arbitrarily in the simulation.

$\lambda_2 = 605$ nm: circle diameter is 13% out of the SLM space size, SLM space size in the simulation is therefore 237 pixels, so zero padding from 215×215 to 237×237 is required.

$\lambda_3 = 705$ nm: circle diameter is 11.5% out of the SLM space size, SLM space size in the simulation is therefore 269 pixels, so zero padding from 215×215 to 269×269 is required.

$\lambda_4 = 800$ nm: circle diameter is 10.1% out of the SLM space size, SLM space size in the simulation is therefore 305 pixels, so zero padding from 215×215 to 305×305 is required.

We use in the simulation 30×30 detector's grid patches in which a random number between [1,2] emitters from each one of the four colors are located in random positions over the 120×120 high resolution grid. Each point is assigned with a random signal value between [15K, 30K photons]. The background term in the Poisson noise is a constant image with a random gray level in the range [144, 676].

The optimized-SLM estimation net was trained for 155K iterations with a learning rate of 0.0001.

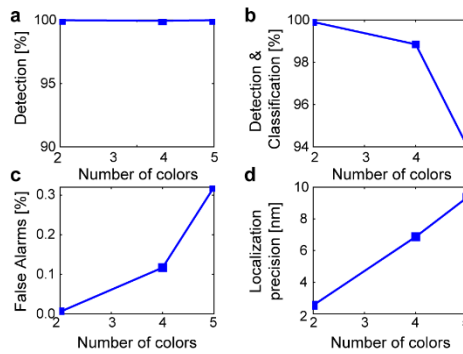


Fig. 17. Evaluating performance over different numbers of colors. (a) Detection performance. (b) Color determination performance. (c) False alarms. (d) Localization performance.

5.5.6. Evaluating performance at extremely-low SNR conditions

In cases when the SNR is very low, the optimized SLM voltage mask only changes the PSF slightly from the normal one (Fig. 18), exhibiting mostly a linear phase ramp resulting in a

non-informative lateral shift, along with a slight non-linear phase pattern. To find the voltage pattern, we simulated 5-10 emitters of each color in a $20 \times 20 \text{ pixel}^2$ FOV, assigning each emitter fed to the physical net with a random signal value between [2400, 4800 photons] (40% of the signal values used in Fig. 9), and added a background term in the range [144, 676 photons].

After training the net for 150K iterations, the received measures are: detection 99%, detection & determination 94.1%, false alarms 0.09% and localization precision 13.5 nm.

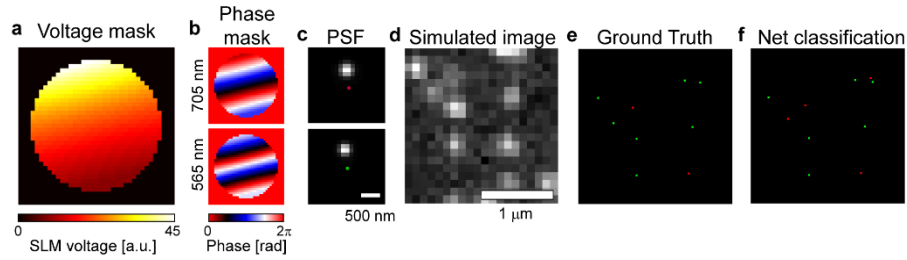


Fig. 18. Optimized SLM results at low-signal conditions. (a) The optimized voltage pattern. (b) The red and green phase patterns. (c) The simulated red and green PSFs. (d) An example of a generated PSFs image with an extremely low SNR. (e) The ground truth of d. (f) Classification by the net.

5.6. Tracking experiments

For tracking experiments, samples were prepared between a glass slide and coverslip, and imaged with the same 4f-imaging system designed for the quantum dot experiments, except that the high-magnification, high NA objective was replaced with a 20X, 0.75 NA lens. This enabled a larger depth of field to accommodate z-movement. One result of the change in objective is a larger back focal plane in our relay system ($\sim 1.13 \text{ cm}$ in diameter, Eq. (4), extending past the size of the liquid crystal SLM in our setup. Despite this constraint, the net converges on an applicable pattern, demonstrating the flexibility of our approach (Fig. 19).

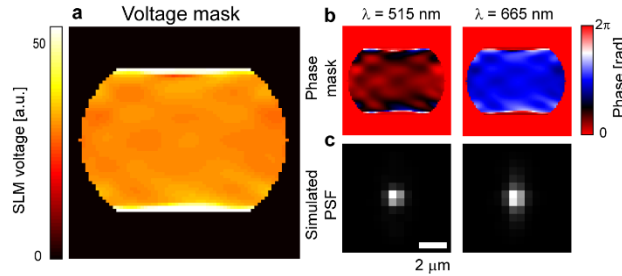


Fig. 19. Optimized SLM pattern for the moving-beads optical system. (a) The optimized voltage pattern. (b) The red and green phase patterns. (c) The simulated red and green PSFs.

To determine if there was an appreciable systematic error caused by using the net-modified PSF, we analyzed the mean squared displacement (MSD) curves obtained for 14 mobile beads, each consisting of 1000 frames (7 red and 7 green). Each trajectory was divided into 10 equal-length segments, and a first order polynomial was fit to the first 5 time lags. For each bead, we performed a two-sample t-test to see if the extracted diffusion coefficients measured with the mask on and off had the same mean. We found there to be no statistically significant change in the measured diffusion coefficient for any beads evaluated ($P_{\text{avg}} = 0.95$, where 0.05 represents a rejection of the null hypothesis), or by other criteria considered, including alternative significance tests (*e.g.* Kolmogorov-Smirnov test) and various maximum time lags used for the MSD curves. The mean and standard deviation of the diffusion coefficients were measured to be 0.45 ± 0.10 and $0.27 \pm 0.10 \mu\text{m}^2/\text{s}$ for the green

and red beads, respectively. The localization precision, which is proportional to the MSD curve offset, did indicate that there was some degradation in the localizability, however. For moving red beads, the calculated precision was 42 nm and 44 nm for the standard and optimized PSFs, respectively; and for green beads the precision was determined to be 42 and 45 nm, with the standard and optimized PSFs, respectively. Analysis of localized static beads stuck to the chamber confirmed a similar change in the precision when using the optimized mask.

5.7. Single-molecule localization microscopy

The performance of the color-identification net depends on the factors described in previous sections. For single-molecule localization microscopy (SMLM), the limited number of useful emitters and limited SNR poses a challenge. To quantitatively assess the applicability of our method to multicolor SMLM, we generated movies of simulated blinking emitters using the Test-STORM toolbox [48] with a known ground truth. In simulations, two fluorophores were simulated: Alexa 647 and Alexa 565 (Figs. 20(a) and (b)). All parameters used for simulations were the defaults of the program, except that for our analysis, two separate movies were combined so that the background was doubled so that the mean background and variance was 400. The reconstruction net was trained using 9915 red and 45215 green emitters over 80K iterations with a learning rate of 0.0001 and batch size 16.

For quantifying the performance, an image was generated to contain spatially-separated fluorophores (Figs. 20(c) and 20(d)). The reconstruction net scores each localization [0,1], which is then used to classify the localizations (Fig. 20(e)). The threshold to classify red and green emitters is a tunable parameter that should be weighted such that the misclassification rate is roughly even. Using a threshold of 0.95, the spatially separated localizations were used to quantify the correct classification rate. For red emitters it was 91% and for green it is 86% (Fig. 20(f)). This threshold was then used for a more qualitative evaluation on the remaining test data sets. Various overlapping structures were simulated (Fig. 20(g)). In the histogram, each pixel was colored given a red and green intensity based on the number of localizations of each red/green classified emitter (Fig. 20(h)).

In biological samples, structures can be more heterogeneous. To test the applicability of our method, we examined fluorescently-labeled HeLa cells. Cells were cultured on glass coverslips, fixed with 4% paraformaldehyde, and then the microtubules and mitochondria were fluorescently labeled with Alexa-647 and Alexa-555 *via* α -tubulin and TOMM20-binding antibodies (ab190573, ab214409, both Abcam) (Fig. 20). Immediately before image acquisition, the sample media was replaced with imaging buffer for blinking single molecules as described previously [18]. For Alexa-647, the median signal observed was 5920 photons and for Alexa-555, 6820 photons were measured, where the background standard deviation was 29 photons per pixel. Net training was done using regions of the image containing only one species, as identified by diffraction-limited images, or in single-antibody labeled samples. Only molecules with a peak intensity $8X$ the standard deviation of the background were used for training. A super-resolved image is reconstructed by creating a Gaussian-blurred-2D histogram of the localized positions (Fig. 20(b)), where each bin was colored using the net's classification score of nearby emitters (Fig. 20(c)). Beyond the successful color-determination in the zoomed-in region of Fig. 20(c), the current limitations of our method are also visible. In highly-challenging conditions (*e.g.* the top part of the cell Figs. 20(g) and 20(h)), the color-differentiation did not match the diffraction-limited image. This may be due to higher emitter-density, increased sample thickness, and non-uniform conditions across the FOV.

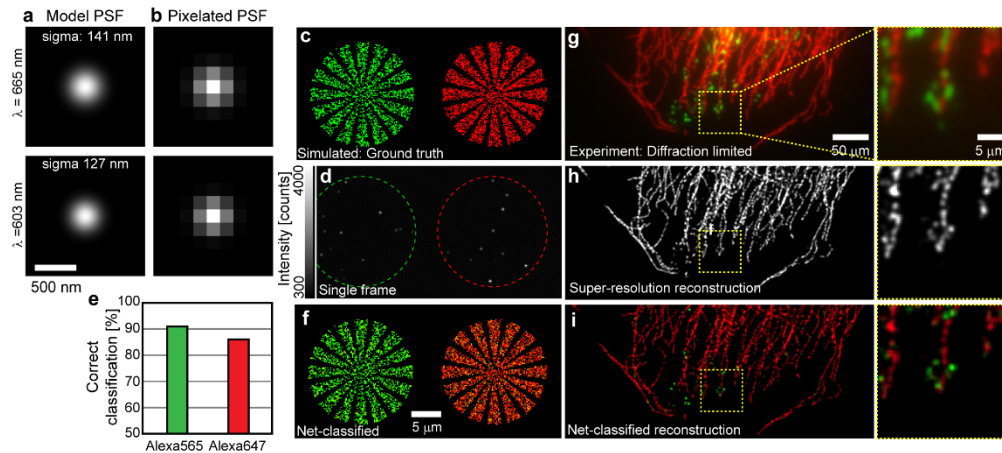


Fig. 20. Single-molecule classification using neural nets. (a) Simulated PSFs for Alexa 647 and Alexa 565. (b) Pixelated PSFs. (c) Ground truth image of two species-specific objects containing 25K fluorophores, each. (d) Representative single frame. (e) Classification success was calculated by the fraction of emitters successfully identified. (f) Pixels colored by the net according to density of emitters. (g-i) A fluorescently labeled HeLa cell. (g) Combined spectrally-filtered, diffraction-limited image. (h) Super-resolution reconstruction. (i) Reconstructed image, with pixels colored according to the net-classifications of individual localizations.

5.8. Microscope characterization

While the diffraction limit depends linearly on wavelength (Fig. 14), experimentally, the shape of the average PSF observed for 565 and 705 nm Qdots was surprisingly similar (Figs. 2(e) and 2(g)). Nonetheless, the net was able to distinguish the two types of emitters (Fig. 2(h)). The likely source of the size-similarity is the imperfect chromatic-correction present in Apo objectives that causes a focal shift between different colors (Fig. 21), as has been observed by others as well [28].

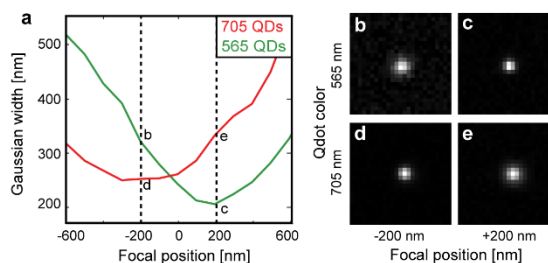


Fig. 21. Chromatic focal shift for a Nikon Apo TIRF 100x/1.49 Oil objective. (a) Focal position versus Gaussian-fit size parameter, sigma. Image data for both Qdots was acquired simultaneously before identifying the color of objects in the FOV by adding a bandpass filter to the image path. (b-e) Intensity-normalized image data showing the standard PSF at two different focal positions for two Qdot emitters.

Funding

H2020 European Research Council Horizon 2020 (802567); Israel Academy of Sciences and Humanities (Alon Fellowship Program); Israel Science Foundation (ISF) (852/17); Ollendorff Foundation; Technion-Israel Institute of Technology (Career Advancement Chairship); Zuckerman Foundation.

References

1. S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.* **91**(11), 4258–4272 (2006).
2. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science* **313**(5793), 1642–1645 (2006).
3. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat. Methods* **3**(10), 793–796 (2006).
4. N. Bourq, C. Mayet, G. Dupuis, T. Barroca, P. Bon, S. Lécart, E. Fort, and S. Lévêque-Fort, "Direct optical nanoscopy with axially localized detection," *Nat. Photonics* **9**(9), 587–593 (2015).
5. T. A. Klar and S. W. Hell, "Subdiffraction resolution in far-field fluorescence microscopy," *Opt. Lett.* **24**(14), 954–956 (1999).
6. T. Dertinger, R. Colyer, G. Iyer, S. Weiss, and J. Enderlein, "Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI)," *Proc. Natl. Acad. Sci. U.S.A.* **106**(52), 22287–22292 (2009).
7. J. Fölling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S. W. Hell, "Fluorescence nanoscopy by ground-state depletion and single-molecule return," *Nat. Methods* **5**(11), 943–945 (2008).
8. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.* **19**(11), 780–782 (1994).
9. G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, and X. Zhuang, "Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging," *Nat. Methods* **8**(12), 1027–1036 (2011).
10. T. Kim, S. Moon, and K. Xu, "Information-rich localization microscopy through machine learning," <https://www.biorxiv.org/content/early/2018/08/07/373878>
11. C. Franke, M. Sauer, and S. van de Linde, "Photometry unlocks 3D information from 2D localization microscopy data," *Nat. Methods* **14**(1), 41–44 (2017).
12. A. P. Bartko and R. M. Dickson, "Imaging three-dimensional single molecule orientations," *J. Phys. Chem. B* **103**(51), 11237–11241 (1999).
13. D. J. Rowland and J. S. Biteen, "Top-hat and asymmetric Gaussian-based fitting functions for quantifying directional single-molecule motion," *ChemPhysChem* **15**(4), 712–720 (2014).
14. M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes," *Science* **317**(5845), 1749–1753 (2007).
15. S. M. Riordan, D. P. Heruth, L. Q. Zhang, and S. Q. Ye, "Application of CRISPR/Cas9 for biomedical discoveries," *Cell Biosci.* **5**(1), 33 (2015).
16. T. Huang, C. Phelps, J. Wang, L. J. Lin, A. Bittel, Z. Scott, S. Jacques, S. L. Gibbs, J. W. Gray, and X. Nan, "Simultaneous multicolor single-molecule tracking with single-laser excitation via spectral imaging," *Biophys. J.* **114**(2), 301–310 (2018).
17. A. von Diezmann, M. Y. Lee, M. D. Lew, and W. E. Moerner, "Correcting field-dependent aberrations with nanoscale accuracy in three-dimensional single-molecule localization microscopy," *Optica* **2**(11), 985–993 (2015).
18. Y. Shechtman, L. E. Weiss, A. S. Backer, M. Y. Lee, and W. E. Moerner, "Multicolour localization microscopy by point-spread-function engineering," *Nat. Photonics* **10**(9), 590–594 (2016).
19. C. Smith, M. Huisman, M. Siemons, D. Grünwald, and S. Stallinga, "Simultaneous measurement of emission color and 3D position of single molecules," *Opt. Express* **24**(5), 4996–5013 (2016).
20. A. S. Backer and W. E. Moerner, "Extending single-molecule microscopy using optical Fourier processing," *J. Phys. Chem. B* **118**(28), 8313–8329 (2014).
21. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
22. A. Chakrabarti, "Learning Sensor Multiplexing Design through Back-propagation," <https://arxiv.org/abs/1605.07078>
23. R. Horstmeyer, R. Y. Chen, B. Kappes, and B. Judkewitz, "Convolutional neural networks that teach microscopes how to image," <https://arxiv.org/abs/1709.07223>
24. S. Elmalem, R. Giryas, and E. Marom, "Learned phase coded aperture for the benefit of depth of field extension," *Opt. Express* **26**(12), 15316–15331 (2018).
25. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458–464 (2018).
26. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," *Nat. Biotechnol.* **36**(5), 460–468 (2018).
27. P. Zhang, S. Liu, A. Chaurasia, D. Ma, M. J. Młodzianoski, E. Culurciello, and F. Huang, "Analyzing complex single-molecule emission patterns with deep learning," *Nat. Methods* **15**(11), 913–916 (2018).
28. C. Cabriel, N. Bourq, P. Jouchet, G. Dupuis, C. Leterrier, A. Baron, M.-A. Badet-Denisot, B. Vauzeilles, E. Fort, and S. Lévêque-Fort, "Combining 3D single molecule localization strategies for reproducible bioimaging," <https://www.biorxiv.org/content/early/2018/11/16/385799>
29. L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
30. A. Jesacher, S. Bernet, and M. Ritsch-Marte, "Colored point spread function engineering for parallel confocal microscopy," *Opt. Express* **24**(24), 27395–27402 (2016).

31. M. Siemons, C. N. Hulleman, R. Ø. Thorsen, C. S. Smith, and S. Stallinga, "High precision wavefront control in point spread function engineering for single emitter localization," *Opt. Express* **26**(7), 8397–8416 (2018).
32. Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, "Optimal point spread function design for 3D imaging," *Phys. Rev. Lett.* **113**(13), 133902 (2014).
33. Y. Shechtman, L. E. Weiss, A. S. Backer, S. J. Sahl, and W. E. Moerner, "Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions," *Nano Lett.* **15**(6), 4194–4199 (2015).
34. W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," <https://arxiv.org/abs/1708.03307>
35. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica* **4**(11), 1437–1443 (2017).
36. P. N. Petrov, Y. Shechtman, and W. E. Moerner, "Measurement-based estimation of global pupil functions in 3D localization microscopy," *Opt. Express* **25**(7), 7945–7959 (2017).
37. L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nat. Methods* **9**(7), 721–723 (2012).
38. S. J. Holden, S. Uphoff, and A. N. Kapanidis, "DAOSTORM: an algorithm for high-density super-resolution microscopy," *Nat. Methods* **8**(4), 279–280 (2011).
39. F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, "Simultaneous multiple-emitter fitting for single molecule super-resolution imaging," *Biomed. Opt. Express* **2**(5), 1377–1393 (2011).
40. S. Gazagnes, E. Soubies, and L. Blanc-Féraud, "High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation," in *IEEE International Symposium on Biomedical Imaging, (ISBI, 2017)*, pp. 28–31.
41. J. Min, C. Vonesch, H. Kirshner, L. Carlini, N. Olivier, S. Holden, S. Manley, J. C. Ye, and M. Unser, "FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data," *Sci. Rep.* **4**(1), 4577 (2015).
42. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods* **9**(7), 676–682 (2012).
43. M. Ovesný, P. Křížek, J. Borkovec, Z. Svindrych, and G. M. Hagen, "ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging," *Bioinformatics* **30**(16), 2389–2390 (2014).
44. A. Vedaldi and K. Lenc, "MatConvNet - Convolutional Neural Networks for MATLAB," in *Proceedings of the 23rd ACM International Conference on Multimedia, (ACM, 2015)*, pp. 689–692.
45. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," <https://arxiv.org/abs/1609.04802>
46. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," <https://arxiv.org/abs/1412.6980>
47. S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **6**(02), 107–116 (1998).
48. T. Novák, T. Gajdos, J. Sinkó, G. Szabó, and M. Erdélyi, "TestSTORM: Versatile simulator software for multimodal super-resolution localization fluorescence microscopy," *Sci. Rep.* **7**(1), 951 (2017).