Published in final edited form as:

Microsc Res Tech. 2018 September ; 81(9): 966–972. doi:10.1002/jemt.23059.

Molecular Imaging with Neural Training of Identification Algorithm

A.J. Nelson¹ and

Department of Physics and Astronomy 5709 Bennett Hall University of Maine, Orono, ME 04469-5709

S.T. Hess*

Department of Physics and Astronomy 5709 Bennett Hall University of Maine, Orono, ME 04469-5709

Abstract

Superresolution localization microscopy strongly relies on robust identification algorithms for accurate reconstruction of the biological systems it is used to measure. The fields of machine learning and computer vision have provided promising solutions for automated object identification, but usually rely on well represented training sets to learn object features. However, using a static training set can result in the learned identification algorithm making mistakes on data that is not well represented by the training set. Here we present a method for training an artificial neural network without providing a training set in advance. This method uses the data to be analyzed, and the fitting algorithm to train an artificial neural network tailored to that data set. We show that the same artificial neural network can learn to identify at least two types of molecular emissions: the regular point spread functions, and the astigmatism point spread function. Simulations indicate that this method can be extremely reliable in extracting molecular emission signatures. Additionally, we implemented the artificial neural network calculation to be performed on a graphics processing unit (GPU) for massively parallelized calculation which drastically reduces the time required for the identification process. By implementing the neural identification on a GPU, we allow this method of identification to be used in a real time analysis algorithm.

GRAPHICAL ABSTRACT

*Corresponding author. samuel.hess@maine.edu.

¹Present Address: Department of Biochemistry, Weill-Cornell Medicine of Cornell University, 1300 York Ave. Box #63 Room A108, New York, NY 10065

AUTHOR CONTRIBUTIONS

A.J.N. designed and analyzed experiments, wrote simulation code, and wrote the manuscript. S.T.H. designed experiments and wrote the manuscript.

Keywords

Super resolution Localization Microscopy; FPALM; PALM; STORM; Deep Learning; Machine Learning; Convolutional Neural Network; Semi-supervised learning

INTRODUCTION

FPALM, PALM, STORM, and other superresolution localization microscopy (SRLM) techniques have helped further our understanding of biology by offering researchers a way to directly watch dynamics at length scales below the optical diffraction limit (200–250nm) (Betzig et al., 2006; Samuel T. Hess, Girirajan, & Mason, 2006; Rust, Bates, & Zhuang, 2006). A key step in the processing of the data for these methods is identifying regions in images that contain fluorescent molecular emissions for subsequent localization. While a multitude of algorithms currently exist to perform this function, usually in tandem with subsequent localization, almost all of these methods rely on some previous knowledge of the data set, and the expected result of molecular emission, to help find the molecules (Izeddin et al., 2012; Olivo-Marin, 2002; Serge, Bertaux, Rigneault, & Marguet, 2008; Smith, Joseph, Rieger, & Lidke, 2010; Smith, Stallina, Lidke, Rieger, & Grunwald, 2015). Furthermore, many of these techniques are not easily generalizable to methods which augment the conventional point spread function (PSF), such as astigmatism, biplane, or double helix PSF imaging. Building an algorithm that can be easily generalized to multiple imaging regimes or methods could allow for a more streamlined flow of data analytics in SRLM experiments.

The field of data science, specifically machine learning and computer vision, may offer insight into this goal. It was shown that multi-layered feedforward neural networks are capable of approximating any function (Hornik, Stinchcombe, & White, 1989). A classic example of the power of these artificial neural networks is an algorithm capable of recognizing numerical digits after learning from a training set (Cheng & Titterton, 1994). Using a basic artificial neural network and a custom feedback learning algorithm, we were able to teach a computer how to identify molecules with the only prior information used being the chosen localization algorithm.

Here we present Molecular Imaging with Neural Training of Identification Algorithm (MINuTIA). This method iteratively trains an artificial neural network on raw data. By basing the success of an identification on how well it was fitted by the localization algorithm, we are able to generate a training set from a subset of raw data frames. This

method has the advantage of automatically selecting against biases that would arise from a user created training set. Additionally, as imaging conditions change from experiment to experiment, this method offers the researcher a way for a tailored identification algorithm despite differences in point spread function shape that would arise from different imaging conditions.

While there is a significant increase in calculation time when compared to some other methods, the neural network identification step was written using CUDA software and is utilized as a parallelized image filter whose calculation is carried out on a mid-range graphics processing unit (GPU). This allows for a massive increase in processing speed which ultimately allows for higher data throughput when compared to simpler serial based methods.

METHODS

We designed an ANN capable of discerning molecular emission images for segmentation and subsequent localization. The implementation of this ANN to the identification problem is essentially one of data classification. Specifically, we attempted to differentiate patterns on 49 pixels into those that contained the signature of molecular emissions, from those that did not. As we attempted to do this mathematically, the problem is analogous to a 2-dimensional problem of determining a linear boundary which separates 2 distinct classes of data.

We decided to train the weighting parameters used by the ANN by implementing the back-propagation algorithm (David E. Rumelhart, Geoffrey E. Hinton, & Ronald J. Williams, 1986) on training data. Initially we used supervised learning techniques by selecting images from experimental data as well as computer generated approximations of molecular emission images. We noticed that the resulting ANN would consistently pick out image regions that resulted in poor localization fits. We realized this was a problem with the images used to train the algorithm and changed our method to an unsupervised learning technique.

The process starts by randomly assigning values to all node weights in the ANN. This randomization helps to break any symmetry artifacts that could arise in the neural network training if all ANN parameters were initialized to the same value. If multiple data sets are being used to train the ANN, first a random data set is chosen, then random frames from the dataset are chosen to be analyzed. The decision was made in an attempt to prevent the training from biasing the final neural net towards any particular data set, or particular frame. The raw frames are preprocessed by using rolling ball subtraction to remove some background from the images (Gould, Verkhusha, & Hess, 2009). The randomized ANN then analyzes every pixel from the background subtracted frames to determine if a molecule is contained in the region surrounding a central pixel. The “identified” areas are segmented and fed into a localization algorithm. Here we used Maximum Likelihood Estimation (MLE) to fit a Gaussian approximation. The results of the fit are then analyzed and subjected to a tolerancing function. This step decides whether an identified region results in a good or poor fit based on user selected tolerances of the fitting parameters and reported errors from the MLE fit.

To help speed the learning process, if an insufficient number of well-fitted regions are identified on the first iteration, a new neural net is generated as described above and new regions are identified and tested. All the regions identified are held in memory. When a sufficient number of positive regions have been found, this step terminates, and the algorithm trains the neural network based on the training set. Without this step, MINuTIA is susceptible to only learning from a collection of negative images, which prevents any subsequent identification of positive regions, and requires the user to restart the process.

The learning algorithm then saves all segmented images and creates an ‘answer key’ consisting of a value of 1 or 0 depending on whether the image (subregion) resulted in a good or poor fit respectively. The segmented images and answer key are used to train the weighting parameters of the ANN by minimizing the following logarithmic cost function.

$$c = \left(\frac{1}{m}\right) \sum_{i=1}^m y_i * \ln(a_i) + (1 - y_i) * \ln(1 - a_i) + \frac{\lambda}{2m} * \sum_{j=1}^n \Theta_j^2 \quad (1)$$

Where C is the final cost, m is the number of training sets, y_i is the desired determination of training set i , a_i is the obtained result using current fitting parameters represented by Θ , λ is an overfitting parameter. The second sum acts to penalize the cost when more fitting parameters are used. The process then repeats. Each iteration adds to the total number of training images. This helps prevent the ANN from repeating mistakes that could occur from removing older images. In the interest of computer memory we removed a random 5% of the positively identified images. This prevented the process from repeating mistakes by removing from the negatively identified images, and makes the process less susceptible to falsely identified positive results. The program is allowed to continue using a new data set and a new list of frames until the ANN identifies successfully fitted image regions at a user determined rate, or until terminated by the user.

We noticed that during the early iterations, the parameters were not optimized enough for high specificity in region selection. This caused a large number of regions to be identified for fitting and subsequent training. To help speed up the learning process we limited the number of frames analyzed in the early iterations to only 10 frames of a data set; after the 15th iteration, this number was increased to 1000 frames from a dataset.

The ANN was organized in a single layered feed forward architecture (Hornik et al., 1989) comprised of an input layer, single hidden layer, and output layer as represented in figure 2. The input layer consists of a linearized array of pixel values which comprise the $n \times n$ pixel region. The activation function for each node in the hidden layer and output layer was chosen to be the Sigmoid function (Cybenko, 1989). The final value of the central pixel of an n by n pixel image is calculated using the following equations.

$$Z_m = \sum_{i=0}^{n^2} \Theta_{mi} * P_i \quad (2)$$

$$a_m = \left(1 + e^{-z_m}\right)^{-1} \quad (3)$$

$$Z_{out} = \sum_{j=0}^m \Theta_j * a_j \quad (4)$$

$$a_{out} = \left(1 + e^{-z_{out}}\right)^{-1} \quad (5)$$

Where P_i is the pixel value of the i th pixel, Θ_{mi} and Θ_{out} are the learned parameters linking the i th pixel to the m th neural node, and the $m+1$ neural nodes to the output node respectively, Z_m and Z_{out} are the arguments to the sigmoid function for the m th neural node and the output node respectively, a_m and a_{out} are the activation values of the neural nodes, a_{out} is the final value output by the neural network. All sums start at the 0 index to incorporate the bias unit needed for proper calculation (D. E. Rumelhart, G. E. Hinton, & R. J. Williams, 1986).

As with most classification and fitting problems, one must determine the optimal number of parameters for use in the fit. Too few parameters result in poor classifications and too many can result in overfitting and inefficiency. In the case of ANNs this choice of parameters is encoded in the number of neural nodes used. We trained ANNs with different numbers of hidden layer nodes to check for underfitting by looking as how well the neural net identified known images. Our training set was divided randomly into a 70 – 30 split where the ANN was trained on 70% of known images and then tested on the remaining 30%. In artificial neural networks, overfitting occurs when the network has memorized the test set; as such, using a neural network with 30 hidden layer nodes would not be susceptible to overfitting if presented with an order of magnitude more test images as was the case here. Instead, increasing the number of nodes would allow the neural network to better identify regions (Hornik et al., 1989) at the expense of computational cost. As such we ran simulations to find that using more than 30 hidden layer nodes did not greatly affect the performance of the neural network (Figure 3).

Each of the pixel values being analyzed is sent to each of the 30 nodes with different weighting parameters. An activation value is calculated for each node which is then sent to the output node. The output node computes the final activation value for the area by again combining the inputs with respective weighting values and computing an activation value that lies between 0 and 1. The decision to segment and localize is made if the output value is above 0. The cutoff of 0.5 is chosen because this is the resulting value of the sigmoid function when the argument is 0. To prevent over counting of a region that has multiple pixels with activation values above 0.5, we add the requirement that the activation value of a pixel must also be the maximal value in a 5 pixel by 5 pixel neighborhood.

The ANN analyzes data around a central pixel to determine an activation value for that pixel. For our implementation we used a 7 pixel by 7 pixel window, but this size can be scaled to accommodate any imaging modality. The activation value represents a measure of certainty that the surrounding region will result in a good fit. The activation value of one pixel is independent of the activation value of any other pixel as the activation values do not depend on each other. This has the advantage of making the problem of measuring an activation value of all pixels inherently parallelizable, and the solution can then be implemented by a massively parallelized approach on a graphical processing unit (GPU). Although the time required to measure the activation value of one pixel may be much longer than for other methods, parallelizing the problem accelerates the analysis by over an order of magnitude (overall), making the analysis of large data sets much more temporally economical on the GPU. This also has the advantage of using previously trained ANN parameters to get a real time estimation of data when this method is paired with a GPU based localization algorithm such as MLE (Smith et al., 2010).

A simulation was made to benchmark MINuTIA's performance. The simulation measured the distributions of experimental parameters images acquired from single color FPALM experiments fitted independently using a published localization analysis method (S. T. Hess et al., 2007). Using these distributions (Figure 4), the simulation produced tens of thousands of simulated molecular images (Figure 5) distributed over ten thousand frames. The molecular images were simulated using the Gaussian PSF approximation whose height, width, position, and offset parameters were chosen to match simulation distributions. Different relative levels of heights and offsets were used to further test the limits of MINuTIA. The same neural architecture was used to analyze all the data sets. After identification, the regions were localized and the results were subjected to goodness of fit tolerances then compared to the known truths of the data sets.

RESULTS

MINuTIA was performed on simulated files (Figure 5). The training was performed over multiple sets of random frames from the file. The results of the full localization analysis with each iteration of learned neural parameters were examined. We noticed that the detection accuracy (the percentage of identifications that match known truths before tolerances) rose very quickly to slightly above 95% in the non-astigmatism sets. False positives, defined as the percentage of false identifications compared to all identifications, quickly fell to values between 6–20%. After tolerances, the detection efficiency dropped to slightly above 90%

while the false positive rate fell below 1%. After 18 iterations, the detection accuracy and false positive rates had stabilized to above 90% and below 0.6% respectively. For the astigmatism data set, the detection efficiency after tolerances was reduced to 80%, but the false positive rate stayed below 1% (Figure 5).

Of the molecules that were not detected there was a noticeable difference between the missed molecules and the ones detected. On average the number of expected photons was much lower for the missed molecules at 170 average photons than for the detected molecules at 230 photons on average. The average ratio of expected photons to the square root of the offset for the missing molecules was also lower at 69 than the average for the detected molecules at 73. This information indicates that the molecules that are missed are the ones that would result in poorer fits, and would be harder to confirm as molecules.

MINuTIA was again trained over multiple data sets containing different molecular brightness values (the total number of photons emitted from a molecule per camera frame). As the average brightness increases, detection accuracy remained above 90% while the false positive rate remains below 1%. There was no obvious advantage to using a training set comprised of homogenous molecules, in comparison to a training set that contained more heterogeneously distributed molecules. The final results for detection accuracy of the dimmest data set was still above 90% with less than 1% false positive identification (Figure 5A).

DISCUSSION

Localization microscopy requires reliable algorithms for molecular identification. We have invented a method for molecular identification which reduces the rates of false positives while maintaining >90% of known molecules. We can think of all the image pixels as mutually orthogonal unit vectors, and the pixel value as a linear multiplier (i.e. component in a particular unit vector direction). Thus, for analysis of super resolution imaging it is essential that we stay within the range of linear response of each pixel to photons. The task we then attempt to accomplish is to define a multivariate function capable of mapping a high dimensionality vector space (the space defined by all possible images) onto a binary output (the decision of whether the image is or is not a fluorescent molecule). Based on the number of possible values for each of the n pixels (typically numbering $n=49$ in our case, each with value between zero and a few thousands of photons), the search is occurring within an n -dimensional hyperspace consisting of more than 10^{236} possible states. Therefore, while the true mapping function may be impossible to know, multi-layered feed forward neural networks are well suited for approximating this function arbitrarily well (Hornik et al., 1989).

Our simulations suggest that MINuTIA can be highly effective at selecting regions containing molecules while avoiding false positives, even in cases where the point spread function has been altered from its normal shape, such as with astigmatic detection. This is imperative for maximizing data fidelity, as well as minimizing analysis time. While the results of our simulations were highly promising, we can only use them as a prediction of the performance of MINuTIA on actual data, namely that MINuTIA is expected to be highly

effective at learning to identify regions in an experimental data set that will result in a successful localization fit.

In our simulations we see that MINuTIA can identify regions for fitting with up to a 90% success rate; however, in actual data we usually see success rates (number of regions resulting in a successful fit compared to total number of regions identified) between 70% and 80%. We attribute this, in part, to our simulations not perfectly representing data acquisition, specifically the contribution of background to the data, the effect of molecular orientation on the final PSF including its variation in shape (Backlund, Lew, Backer, Sahl, & Moerner, 2014) and intensity (Valles & Hess, 2017), and defocusing of molecules near the focal plane. The success rate in the experimental data sets represents the percentage of regions that result in a good fit. Without additional information, and a far more elaborate experimental procedure, it is impossible to know the actual detection efficiency of an analyzed experimental data set.

The results from our simulations are an improvement in detection accuracy over other recently announced techniques, while maintaining similar false positive rates (Smith et al., 2015). However unlike other techniques, we make very few assumptions about the point spread function. The significantly reduced number of assumed parameters allows for our technique to fit both astigmatism and unaltered PSFs without the need for adjustment to the algorithm. As such this technique can be thought of as a generic molecular emission identification algorithm potentially compatible with any localization microscopy experiment that utilizes a fitting algorithm. Simply by feeding the output of the fitting algorithm into MINuTIA will allow for the training and learning of an image that would result in a successful fit.

We notice there is fluctuation in the success rate as iterations continue to increase. As the process is allowed to continue, we see that the success rate plateaus within a window generally between 70 and 85% successful identification. We attribute this to the algorithm using different files and frames for each iteration. We make this identification based on the fact that we see significantly less fluctuation in the success rate on simulated data, and a much tighter plateau window. However on actual data there is much more cell to cell variability in key parameters, which can have an impact on imaging conditions in ways unaccounted for by the simulation. As background levels, fluorescent protein expression, and other variables fluctuate between cells, we expect to have some data sets with low background and easily localizable molecules, and some data sets with higher background and molecules that are more difficult to localize. We expect that this variation in data set conditions leads to the fluctuation in success rate seen with experimental data. To a much lesser extent there is some variation in success rate that arrives from how the training set images are handled. To avoid overloading the computer memory, we remove a percentage of images at random after each iteration. Without this step, the learning becomes progressively, and eventually prohibitively, slow. We have still seen the fluctuation in success rate when we do not remove any images from the training set, but we acknowledge that by removing examples of 'good' and 'bad' molecules, the algorithm is altering the training, and thereby the success rate.

We noticed that within the first couple of iterations, MINuTIA becomes increasingly efficient at identifying molecules. After the 15th iteration, the efficiency suffers a bit. On this iteration we increase from 100 to 1000 frames. This is done to prevent the early iterations of the neural net from choosing too many non-specific regions which would make the process much slower than necessary. After this point we notice that MINuTIA quickly asymptotes to some value for percentage of successful fits of the total identifications. We also notice that the detection accuracy also asymptotes to some value above 80% for our simulations. The false positive rate also asymptotes much more quickly and stays consistently low.

The MINuTIA algorithm has been demonstrated with a Gaussian fitting algorithm to learn what images will result in a 'good' fit and discriminate against regions that would not. However, this fitting algorithm is not necessary for the MINuTIA algorithm to perform, and in principle MINuTIA could be paired with any fitting algorithm to learn to differentiate regions that would result in good fits from those that would not. As such we posit that the MINuTIA algorithm is adaptable to any fitting function. This should allow MINuTIA to be incorporated into other localization experiments such as biplane (Juetten et al., 2008), or experiments that involve point spread function shaping (Pavani et al., 2009). To help speed up learning, test images of Gaussians were provided to MINuTIA early on in the process; analogously, test images of the expected PSF could be provided for experiments where the PSF deviates significantly from a Gaussian approximation. The MINuTIA algorithm could potentially be used for analysis of localized data. The algorithm described uses assumptions about structure to speed up learning, but could be adapted to recognize any spatial pattern if MINuTIA is provided with a way to determine the success of a region.

SUMMARY AND CONCLUSION

Here we describe an artificial neural network which can learn the spatial pattern of a pixelated point spread function with a high degree of accuracy. This method allows for reduced user-induced bias in the identification process by allowing the MINuTIA algorithm to decide which spatial patterns are important for identification. This process also has an advantage over supervised artificial neural networks in that it can learn from its mistakes, which supervised neural networks cannot do as easily. Simulations show that this process is able to identify different patterns with no change to neural architecture. This suggests the MINuTIA process is a generalized molecular identification algorithm that could be used in experiments where manipulation of the point spread function is performed such as biplane or double-helix PSF imaging.

This technique further investigates the uses of artificial neural networks in localization microscopy, which help to combine a very powerful data science technique with a technique capable of producing massive amounts of data. Improvements to this technique could arise from investigating different neural network architectures including adding hidden layers, or convolutional layers, to the process, at the cost of development, learning, and computational time. Additionally this technique could potentially be paired with other localization algorithms such as multi-emitter fitting (Huang, Schwartz, Byars, & Lidke, 2011).

Using a robust identification algorithm could allow for the detection of molecules with lower signal to noise ratios, such as those seen in high speed imaging (Nelson, Gunewardene, & Hess, 2014). Incorporating this technique into high speed experimental designs could allow for a more reliable final render using fewer acquisition frames than what has been previously presented. While this technique itself helps to deal with the issue of increased background and lowered signal, pairing it with a high density localization algorithm such as multi-emitter fitting (Huang et al., 2011) could allow for improved sampling on similar time scales, thus improving the overall spatial (or temporal) resolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors thank Brady Butler, Matthew Parent, and Michael Mlodzianoski for useful discussions, and Patricia Byard for technical assistance. This work was funded by National Institutes of Health grants R15GM116002 and R15GM094713, by Maine Technology Institute grants MTAF 1106 and 2061, by the University of Maine Office of the Vice President for Research, and by the Maine Economic Improvement Fund. STH holds licensed patents related to localization microscopy.

REFERENCES

- Backlund MP, Lew MD, Backer AS, Sahl SJ, & Moerner WE (2014). The role of molecular dipole orientation in single-molecule fluorescence microscopy and implications for super-resolution imaging. *Chemphyschem*, 15(4), 587–599. doi: 10.1002/cphc.201300880 [PubMed: 24382708]
- Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, . . . Hess HF (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793), 1642. [PubMed: 16902090]
- Cheng B, & Titterton DM (1994). Neural Networks: A Review from a Statistical Perspective 2–30. doi: 10.1214/ss/1177010638
- Cybenko G (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303–314. doi: 10.1007/BF02551274
- Gould TJ, Verkhusha VV, & Hess ST (2009). Imaging biological structures with fluorescence photoactivation localization microscopy. *Nat. Protocols*, 4(3), 291–308. [PubMed: 19214181]
- Hess ST, Girirajan TPK, & Mason MD (2006). Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophysical Journal*, 91(11), 4258–4272. doi: 10.1529/biophysj.106.091116 [PubMed: 16980368]
- Hess ST, Gould TJ, Gudheti MV, Maas SA, Mills KD, & Zimmerberg J (2007). Dynamic clustered distribution of hemagglutinin resolved at 40 nm in living cell membranes discriminates between raft theories. *Proc Natl Acad Sci U S A*, 104(44), 17370–17375. [PubMed: 17959773]
- Hornik K, Stinchcombe M, & White H (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. doi: http://dx.doi.org/10.1016/0893-6080(89)90020-8
- Huang F, Schwartz SL, Byars JM, & Lidke KA (2011). Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomed Opt Express*, 2, 1377–1393. doi: 10.1364/BOE.2.001377 [PubMed: 21559149]
- Izeddin I, Boulanger J, Racine V, Specht CG, Kechkar A, Nair D, . . . Sibarita JB (2012). Wavelet analysis for single molecule localization microscopy. *Optics Express*, 20(3), 2081–2095. doi: 10.1364/OE.20.002081 [PubMed: 22330449]
- Juette MF, Gould TJ, Lessard MD, Mlodzianoski MJ, Nagpure BS, Bennett BT, . . . Bewersdorf J (2008). Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nature Methods*, 5(6), 527–529. [PubMed: 18469823]

- Nelson AJ, Gunewardene MS, & Hess ST (2014). High speed fluorescence photoactivation localization microscopy imaging
- Olivo-Marin J-C (2002). Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9), 1989–1996. doi: [http://dx.doi.org/10.1016/S0031-3203\(01\)00127-3](http://dx.doi.org/10.1016/S0031-3203(01)00127-3)
- Pavani SRP, Thompson MA, Biteen JS, Lord SJ, Liu N, Twieg RJ, . . . Moerner WE (2009). Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc Natl Acad Sci U S A*, 106(9), 2995–2999. [PubMed: 19211795]
- Rumelhart DE, Hinton GE, & Williams RJ (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Rumelhart DE, Hinton GE, & Williams RJ (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1).
- Rust MJ, Bates M, & Zhuang X (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Meth*, 3(10), 793–796. doi: http://www.nature.com/nmeth/journal/v3/n10/supinfo/nmeth929_S1.html
- Serge A, Bertaux N, Rigneault H, & Marguet D (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat Meth*, 5(8), 687–694. doi: http://www.nature.com/nmeth/journal/v5/n8/supinfo/nmeth.1233_S1.html
- Smith CS, Joseph N, Rieger B, & Lidke KA (2010). Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat Meth*, 7(5), 373–375. doi: http://www.nature.com/nmeth/journal/v7/n5/supinfo/nmeth.1449_S1.html
- Smith CS, Stallinga S, Lidke KA, Rieger B, & Grunwald D (2015). Probability-based particle detection that enables threshold-free and robust in vivo single-molecule tracking. *Molecular Biology of the Cell*, 26(22), 4057–4062. doi: 10.1091/mbc.E15-06-0448 [PubMed: 26424801]
- Valles M, & Hess ST (2017). A Cross Beam Excitation Geometry for Localization Microscopy. *iScience Notes* doi: <http://doi.org/10.22580/2016/iSciNoteJ2.2.1>

RESEARCH HIGHLIGHTS

Here we present a machine learning algorithm for identifying point-spread functions without the need for an a priori training set. We show that this method can detect over 90% of molecules with less than 1% false positive identification in simulations. We further show that because this algorithm does not make assumptions of about the shape of molecular emission, it is compatible with models beyond the symmetric 2D Gaussian.

Figure 1.

Results of the training process are shown above. (A) Identifications that result from the untrained neural net. (B) Identification result after a single training session. (C) Identification result after many training sessions. Note the increasingly preferential nature of the neural network to identify regions that will result in a strong fit.

Figure 2. General architecture of an artificial neural network using a single hidden layer. Pixel information is fed into each hidden neural node and the resulting values from each output node is fed into the output node. For both the input layer and the hidden layer the first value is a bias value fixed to 1.

Figure 3.

Results of identification accuracy on known data. The data was divided into two parts with 70% being used to train the neural network and 30% being used to test how well it performed on new data. Above 30 nodes in size, the improvement of the performance of the network is greatly reduced, indicating a natural criterion for selection of node number.

Figure 4. Representation of simulated data for different numbers of photons detected and different background levels. (A) Low signal to noise example. (B) Mid-range signal to noise example. (C) High signal to noise ratio example. (D) Example of a data set with astigmatic PSF.

Figure 5.

Results of the analyzed simulation data sets after tolerances have been applied as learning is iterated. (A) Results of the low signal to noise set (B) Results of the medium signal to noise set (C) Results of the high signal to noise set. See also Figure 4. In these cases, the rate of molecular identification plateaus at ~90% while the rates of false positive identification fall below 1%. (D) Results of learning on astigmatic PSF data. Here the rate of molecular identification plateaus at just below 80% while the rate of false positives remains below 1%.