Master Thesis

Analyzing single molecule emission patterns using Deep Learning

Anish Mukherjee

Technische Universiteit Delft

TU Delft

Master Thesis

Analyzing single molecule emission patterns using Deep Learning

by

Anish Mukherjee

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday August 25, 2020 at 9:30 AM.

Student number:	4824849	
Project duration:	December 17, 2019 – August 25, 2020	
Thesis committee:	Prof. Dr. S. Stallinga,	TU Delft, Chair and Supervisor
	Prof. Dr. B. Rieger,	TU Delft, Thesis Committee Member
	Dr. Z. Perko,	TU Delft, Thesis Committee Member
	M.Sc. R.O Thorsen	TU Delft, PhD Supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

This master thesis marks the culmination of my two-year long journey of learning at TU Delft. This place will forever hold a very special place in my heart as I got to pursue my lifelong dream of doing cutting-edge scientific research. At TU Delft, I got to pursue a very unique curriculum which was a beautiful amalgamation of the three things I am very passionate about which are Nuclear Physics, Imaging Physics and Artificial Intelligence. During the last year, I got a chance to venture into the field of Computational Imaging starting with my internship at DELMIC and then doing research for my master's thesis in the Quantitative Imaging Lab under the supervision of Dr Sjoerd Stallinga and Dr Bernd Rieger.

First and foremost, I would like to thank Dr Sjoerd Stallinga who gave me this incredible opportunity to venture into the world of nanoscopy. This thesis would not have been possible without his constant guidance. His suggestions and critique helped me not only to grow as a graduate student but also I learnt skills which would help me throughout my whole career and my life. My first lecture in Delft was 'Imaging Systems', a class Dr Stallinga taught and I was so fascinated by the things I learnt that I knew at that very instant I would like to direct my academic career towards this field of study. The fundamentals of optical imaging systems learnt during the course were the foundation of this thesis work.

I would also like to thank Dr Bernd Rieger whose inputs during the early stages of this project helped shape the direction of the research. I would also like to extend my gratitude towards my daily supervisor Rasmus Thorsen who was constantly there for me throughout this journey which seemed daunting in the beginning. His critical inputs and the constant feedback helped me refine the quality of this work continuously. I would also like to extend my gratitude to Ronald Ligteringen for his help in getting me started with the high-performance cluster which was essential for this project. It was due to Ronald that when the whole world ground to a halt due to the pandemic this project went on smoothly. I would also like to thank all the wonderful people at DELMIC especially Daan, Sander, Andries and Thomas who gave me a chance to learn about localization microscopy hands on. The skills I learnt and the knowledge I gained while doing the internship was very helpful during my master thesis. I would also like to thank Teun, Maarten and Thijs with whom I had a memorable time discussing our work.

Coming to the Netherlands and studying in Delft was a dream which I had for a long time. My parents shared this dream of mine and it is only because of them I could fulfil this dream. They were with me every step of the way and they lived this dream of mine vicariously through me when I told them about all the wonderful things I saw and experienced here. I am eternally grateful for all the unconditional love and affection and I hope to make them proud one day by being a good son to them. I would also like to thank my cousin Amitabh and sister in law Urvashi for being a constant source of guidance and inspiration.

It is rightly said that friends are the people who make us smile brighter and live better. I would like to mention a few people who make my life better. My beautiful girlfriend, Komal, who despite being a world away always felt right next to me. She inspired me every step of the way and believed in me throughout the whole journey. For me, she is the embodiment of perfection, courage and diligence and I am grateful to her that she shared this beautiful journey with me. To my best friend in the whole world, Aayush, who is always with me through thick and thin, thank you for being the brother I never had. I would also take a chance to thank my 'Boscoites' who turned into a family from being friends.

I would like to dedicate this thesis to my friends and family because of whom I was able to come here and chase my dreams.

Anish Mukherjee
Delft, August 2020

Abstract

The time taken to generate a super-resolution image and the quality of the final synthetic image depends on the performance of the localization algorithm which is used in the localization microscopy pipeline. The most precise and accurate algorithms are mostly iterative and they take a long time to generate the localization list while the faster 'one-shot algorithms' are not very accurate and precise. A deep learning method smNet (single-molecule Net) was developed by Zhang *et al* [76] which was claimed to perform one-shot localization with precision close to the theoretical limit and very accurately, along with performing aberration estimation and dipole-emitter orientation angle estimation. The deep learning model smNet was trained either by augmenting experimental data or using simulated data generated with an erroneously simplified simulation model and a phase retrieval method. The purpose of this work was to characterize the performance of smNet when it was trained with simulated images generated using an accurate vector model for a range of physical conditions. Along with the characterization of smNet's performance in doing 3D localization and aberration estimation with the accurate vector model, a pipeline was also designed which made the training process of smNet more efficient and computationally cheaper while performing accurate and precise 3D localization and aberration estimation. The pipeline was designed to implement the concept of simulator learning where a smNet model could be trained on simulated data and used to perform 3D localization and aberration estimation directly on experimental data without any retraining or domain adaptation techniques.

Keywords : Localization Microscopy, Deep Learning, 3D Localization, Aberration Estimation

Contents

List of Figures

List of Tables

1 Introduction

1.1. Localization Microscopy

Today, fluorescence microscopy has become one of the most important tools for researchers to see biological structures present inside a cell with great detail. The rise in the use of fluorescence microscopy by researchers can be attributed to the high contrast in the images as a result of fluorescent labels and the ability of the researchers to label the relevant bio-molecules such as proteins with these labels. The introduction of the green fluorescent protein (GFP) in the 1990s and the ability of the researchers to genetically engineer cells to express this protein made fluorescent microscopy a widely used tool in cell research [47]. Figure 1.1 shows the schematic of a fluorescence microscope. Laser light of a certain wavelength (λ_{exc}) is directed through a dichroic mirror and an objective lens to shine on the biological sample. This excitation beam is absorbed by the biological sample and typically a few nanoseconds later, the sample emits a light (λ_{emit}) which is Stokes-shifted by $\sim 10 - 100nm$ to a longer wavelength. This light is captured by the objective lens and passes through a series of dichroic mirror and emission filter to eliminate the reflected excitation beam. The fluorescence emission passes through a tube lens and is focused on to the sensor of a CCD or CMOS camera which generates the image of the biological sample.

The resolution of a state of the art fluorescence microscope is limited by the Abbe's diffraction limit. According to this limit the resolution that can be achieved by an optical microscope is $\lambda/2NA$ where λ is the wavelength of the light used for imaging and NA is the numerical aperture of the objective lens. The typical resolution which is achieved by a modern fluorescence microscope using visible light and a high NA lens is $\sim 200nm$ which limits the imaging of the finer details of structures within the cell. Electron microscopes provide a much higher resolution but they are not useful for live-cell imaging nor it is possible to label specific proteins and hence it does not provide a complete picture of the biological processes going on inside the cell. Over the last few decades, researchers have tried to bypass Abbe's diffraction limit using techniques known as super-resolution techniques. The super-resolution techniques such as Stochastic Optical Reconstruction Microscopy (STORM) and Photo Activated Localization Microscopy (PALM) belong to the field of optical nanoscopy [25][3][62][31]. Using the conventional microscope setup they offer a resolution of $\sim 20nm$ which is about ten times better than the diffraction limit. The diffraction limit is passed by a technique called localization microscopy. In localization microscopy, in each frame, a small number of stochastically switched-on fluorophores are imaged. The centre of each airy spot is localized and using many such frames a super-resolved image is reconstructed using the localization information. Figure 1.2 shows the process of capturing multiple frames with a stochastic fraction of 'on' fluorophores which are used to generate a super-resolution image. This increment in the resolution does not require any additional hardware. The state of the art localization microscopy techniques requires only a conventional fluorescence microscope, laser light sources and a CCD or CMOS camera with low readout noise and a high efficiency of converting photons into electrons. The only additional requirement is the image processing software which processes each frame, generates localization information and reconstruct the super-resolved image from the data.

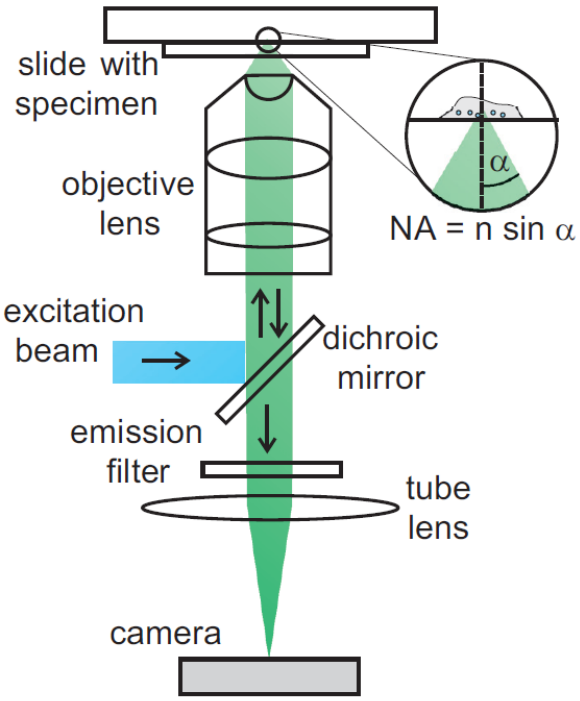


Figure 1.1: Schematic representation of the fluorescence microscopy setup [47].

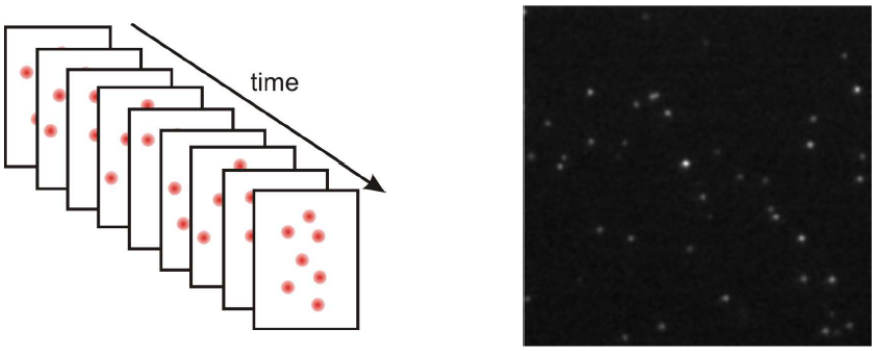


Figure 1.2: Generation of individual frames with a fraction of 'on' fluorophores [62].

1.2. Image Processing in Localization Microscopy

The most important tool to generate a super-resolution image is the image processing software. The performance of the image processing software determines the quality of the output image. Figure 1.3 shows the data flow in an image processing software to generate a super-resolution image. The first task performed by the software is to segment the region of interest (ROI) containing a candidate emitter. The most basic algorithm which is used for segmentation is thresholding [25][3]. The selection of a candidate emitter is done by identifying a pixel which has intensity higher than a defined threshold or intensity more than a fixed multiple of the background. These pixels are used to generate ROIs from a frame with these pixels being the centre of a region of interest. Apart from simple thresholding, more complex wavelet-based algorithms have been proposed. These decompose a raw image into wavelet maps which are used to separate blobs from noise and background and then further watershed segmentation is used to identify the candidate pixel and generate ROIs [51][29]. Another advanced segmentation approach is by local hypothesis testing against the null hypothesis that a pixel belongs to the local background [35]. These techniques have an underlying assumption that the background is uniform and this assumption holds for a small ROI. For cases where the assumption doesn't hold

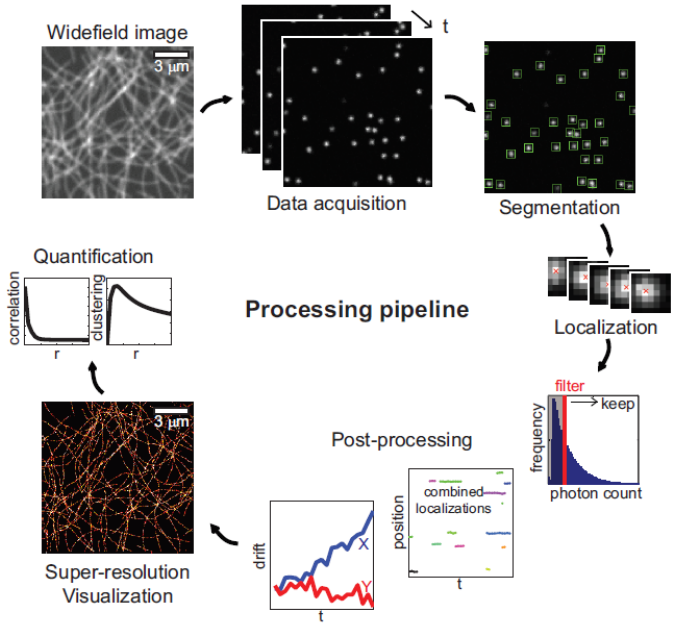


Figure 1.3: Flowchart showing the various stages required in order to generate a super resolution image [47].

median filtering has been proposed to identify candidate emitter pixels [24].

The next step is to localize the centres of the candidate emitters with sub-pixel accuracy. Localization algorithms are discussed in chapter 2 in detail. The most commonly used algorithms are fitting based algorithm. In fitting based algorithms, a point-spread function (PSF) model is fitted to the candidate emitters and the information about the centre of the emitter is obtained. The fitting based algorithms are generally very computationally expensive as they are iterative algorithms. Deep learning-based algorithms are emerging in the field of localization microscopy.

Once localization is done, three important post-processing steps are followed before generating a super-resolution image. The first post-processing step is to eliminate unreliable localizations. This filtering step is performed to remove localizations arising from the emission of multiple fluorophores or localization due to sample contamination or due to autofluorescence. This filtering is done based on the information returned from the localization algorithm. Information about the width of the bead, the intensity of the bead, goodness of fit when a fitting based localization algorithm is typically combined to filter out unreliable localizations [31]. The second post-processing step is to combine localization of the same emitter across the subsequent frames where the emitter is on. Ideally, a single emitter should appear only in one frame but often the same emitter is present in multiple frames. These localizations should be combined based on proximity to generate an accurate representation of the underlying structure. The third post-processing step is to compensate for drift. A typical experiment can last anywhere from a few minutes to a few hours. Over this duration due to mechanical vibrations in the setup and from the surrounding environment can cause the sample to drift relative to the camera setup over a distance which is larger than the localization precision of the algorithm. Drift can be fixed by using mechanical solutions which actively compensate for mechanical vibrations present in the system or the surrounding [25]. Drift correction can be done using algorithms by introducing fiduciary markers in the images and then localizing the position of the fiduciary markers and comparing its position in the subsequent frames using cross-correlation and then compensating for the shift [54][43][19].

The final step of localization microscopy is to generate a super-resolution image. Localization microscopy is inherently different from other microscopy techniques as in localization microscopy, the final result is a set of position estimates while other microscopy techniques generate pixelated images. These positions estimates are used to generate a synthetic image of sub-cellular reality. The simplest visualization technique is to plot the localization estimate as symbols in a Cartesian coordinate system. The resulting Scattergram [3] would depict the underlying structure. Another technique which is used to generate the final visualization is histogram binning where the total field of view is divided into

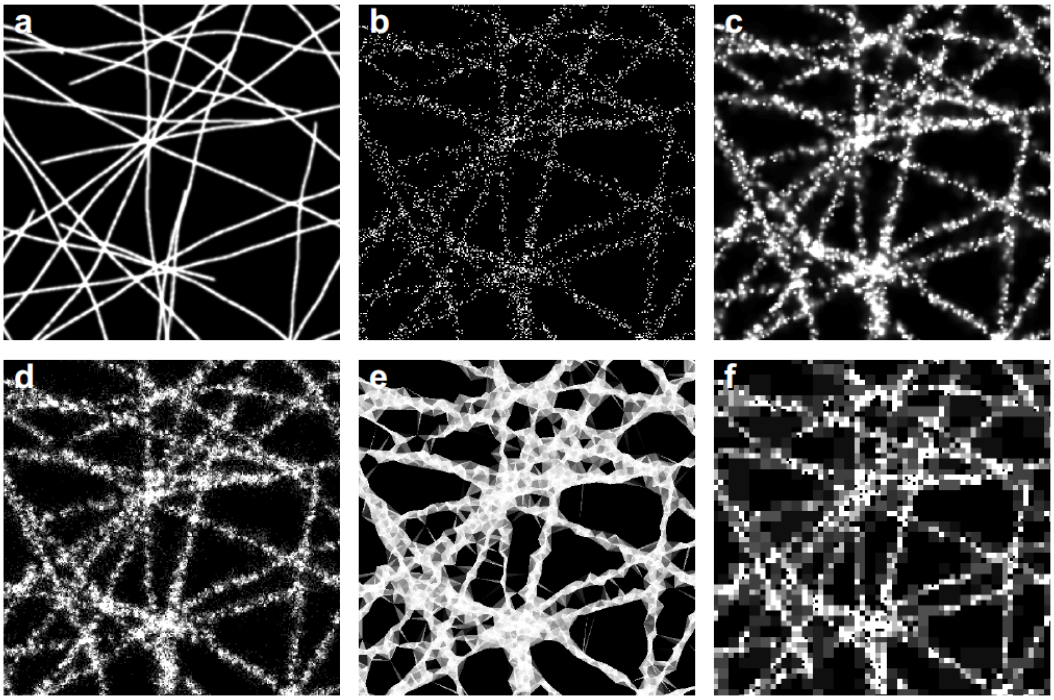


Figure 1.4: Representation of different visualization techniques applied on stimulated localization data of filaments with $\rho = 2.0x10^3 \mu m^{-2}$ and localization precision $\sigma = 10nm$. Panels: (a) The ground truth structure, (b) histogram binning, (c) Gaussian rendering, (d) jittering, (e) Delaunay triangulation, (f) quad-tree visualization [47].

square pixels and each pixel is assigned an intensity value which depends on the number of localizations present inside a pixel [48]. If the size of the pixel is larger than one-quarter of the resolution then the image quality is deteriorated. Images generated by this technique are often jittery because of low SNR. Blurring the image with a radially symmetric kernel improves the image quality [16]. Another technique which is used to generate the super-resolved image is using a Gaussian blob with dimensions depending on the lateral and axial localization precision of each localization [26]. Using the Delaunay triangulation technique, a tiling pattern is created using the localization estimates as the vertices of the triangle. The smaller the triangle the higher is the emitter density and the grayscale intensity assigned to a tile is directly proportional to the emitter density in the final super-resolved image. The quad-tree algorithm generates a visualization by generating square tiles whose size is directly proportional to the emitter density. Initially, each bin is split into 4 sub-region. If the sub-region has more certain a threshold number of emitter then the sub-region is split into 4 more sub-region and the process is repeated until the sub-region cannot be split further [6]. Figure 1.4 shows the resulting super-resolved image generated using different visualization algorithms.

1.3. Research Problem

The fitting based localization algorithms perform localizations with precision close to the theoretical limit (discussed in section 3.3). They have a few drawbacks that since they are iterative algorithms they are computationally expensive and slow. This makes the process of generating a super-resolution image quite long. Deep learning-based methods which are coming up in the field of localization microscopy can perform 3D localization very fast and are computationally inexpensive. Single-molecule net (smNet) [76] is one such deep learning technique which can perform 3D localization. The benefit of using smNet is that along with performing 3D localization it is claimed to also perform angle orientation estimation of a dipole and estimation of wavefront aberrations present in the optical system accurately and precisely. The training process of the smNet algorithm has a few flaws. smNet can be trained either using experimentally obtained data or using simulated data. When smNet is trained using experimentally obtained data, an $N \times N \times L$ stack is created where N is the height and the width of the image and L is the number of experimentally obtained images. The images are converted from a 3D array to a

2D array with a dimension of $N^2 \times L$ and then data augmentation is performed by interpolation along each of the columns and then each row is converted into a $N \times N$ image. This synthetic augmentation technique doesn't accurately represent all the possible variations that could arise from different imaging conditions. When training the smNet with simulated images, the pupil function is extracted from experimental data using a phase retrieval method and then PSFs are simulated at arbitrary positions using the diffraction model and the extracted pupil function. This process has a drawback that the model used to generate the simulated images is an erroneous oversimplified model which doesn't accurately represent the real experimental PSFs and the training process requires extraction of the pupil function from experimental data using a phase retrieval method. The concept of simulator learning is also not tested by the authors where a smNet model trained on purely simulated data can be used to make predictions on real experimental data.

The first challenge was to characterize the performance of smNet trained with simulated images generated using an accurate PSF model over a wide range of physical parameters without the use of any phase retrieval method. Once the characterization was done the next challenge was to design a pipeline which would ensure robust 3D localization using smNet over a large range of physical conditions such as photon count, background count, aberration modes and aberration magnitude. The final challenge was to test the feasibility of the concept of simulator learning using smNet and to compare its performance with a 'conventional' fitting based localization algorithm.

1.4. Thesis Structure

In this section, the structure of the thesis is described. In chapter 2, literature survey is described where localization algorithms are discussed. At the end of the literature survey, some of the unanswered research questions in the localization microscopy are discussed. The physics behind image formation in an optical imaging system along with the concept of the fundamental limit to precision, the so called Cramer Rao Lower Bound is presented in chapter 3. The workflow and description of the building blocks of the smNet neural network are described in chapter 4. Chapter 5 contains information about the simulations carried out. The results obtained from the simulation experiments and experiments performed on empirically collected data are presented and discussed in chapter 6. Inferences, conclusions and recommendations are presented in chapter 7.

2

Literature Survey

In this chapter, various conventional localization microscopy algorithms and their working are discussed. Also, the emergence of deep learning in different stages of localization microscopy are discussed extensively in this chapter.

2.1. Conventional Localization Microscopy Algorithms

In this section, various traditional localization microscopy algorithms are discussed which are not based on the concept of deep learning.

2.1.1. Single Emitter Fitting Based Algorithms

GaussMLE

Smith *et al* [60] came up with an iterative fitting algorithm which converges to the theoretical lower bounds (Cramer-Rao lower bound). This method gives precision which matches the Cramer Rao lower bound over a large range of conditions but this method has a drawback that it's performance degrades when the SNR is very low [2]. Also this method is computationally very expensive as compared to the other methods described in this study. This algorithm computes the maximum likelihood estimate of the particle's position, photon count and the background count using a Newton-Raphson optimization scheme. The PSF model which is used for the algorithm is defined by :

$$PSF(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x-\theta_x)^2-(y-\theta_y)^2}{2\sigma^2}} \tag{2.1}$$

where $\theta_{x,y}$ are the position estimate and accounting for the finite size of the pixel, the imaging model is re-written as :

$$\mu_k(x,y) = \theta_{I_o} \int_{A_k} PSF(u,v)dudv + \theta_{bg} \tag{2.2}$$

and which can be further simplified as :

$$\mu_k(x,y) = \theta_{I_o} \Delta E_x(x,y) \Delta E_y(x,y) + \theta_{bg} \tag{2.3}$$

where $\Delta E_x(x,y)$ and $\Delta E_y(x,y)$ is defined as :-

$$\Delta E_x(x,y) = \frac{1}{2} erf\left(\frac{x-\theta_x+\frac{1}{2}}{2\sigma^2}\right) - \frac{1}{2} erf\left(\frac{x-\theta_x-\frac{1}{2}}{2\sigma^2}\right) \tag{2.4}$$

$$\Delta E_x(x,y) = \frac{1}{2} erf\left(\frac{y-\theta_y+\frac{1}{2}}{2\sigma^2}\right) - \frac{1}{2} erf\left(\frac{y-\theta_y-\frac{1}{2}}{2\sigma^2}\right) \tag{2.5}$$

Using this model the idea is to maximize the function $ln(L(\vec{x}|\theta))$ which is equal to the maximum likelihood estimate of the parameter $\theta_{ML} = arg_{\theta} max L(\vec{x}|\theta)$ where the function $L(\vec{x}|\theta)$, derived by modeling shot noise present in the images as a Poisson distribution, is defined by :

$$L(\vec{x}|\theta) = \prod_{k} \frac{\mu_k(x,y)_k^x e^{-\mu_k(x,y)}}{x_k!} \tag{2.6}$$

and the update rule after each iteration is given by :

$$\theta_i \rightarrow \theta_i + [\sum_{k} \frac{\partial \mu_k(x,y)}{\partial \theta_i}(\frac{x_k}{\mu_k(x,y)} - 1)][\sum_{k} \frac{\partial^2 \mu_k(x,y)}{\partial \theta_i^2}(\frac{x_k}{\mu_k(x,y)} - 1) - \frac{\partial \mu_k(x,y)^2}{\partial \theta_i}]^{-1} \tag{2.7}$$

This method achieves precision which is equal to the information limit or the Cramer-Rao lower bound. Cramer-Rao lower bound can be defined using :

$$\Delta\theta \geq \frac{1}{\sqrt{i(\theta)}} \tag{2.8}$$

where $i(\theta)$ is the information content which is computed from the PSF also known as the Fisher's information matrix, N is the number of photon and θ is the position co-ordinates.

Fitting using C-Spline based Experimental PSF

Li *et al* [36] came up with an MLE based fitting algorithm which uses experimental PSFs instead of the popular Gaussian PSF model. The biggest advantage of this method is this method is compatible with all PSF engineering methods and it can even perform 3D localization without any additional optical component to break the symmetry. This method can utilize the subtle difference in the PSF in both the planes (above and below focus) and use that to perform 3D localization. One major disadvantage of this algorithm that it can only perform single emitter fitting and a high quality model is necessary when experimental PSFs are used.The algorithm works by segmenting ROIs by finding local maxima to identify the candidate molecule followed by sub pixel alignment. Sub-pixel alignment is done by performing 3D cross correlation and the central part of the cross correlation is zoomed by a factor of 20 using c-spline interpolation and the sub-pixel x,y and z shifts are determined. This is followed by shifting the bead with c-spline interpolation. The bead stack is further regularized by performing smoothing with the help of smoothing b-spline in z direction. The smoothened bead stack is up-sampled to get the c-spline co-efficient and then MLE is performed with the help of the Levenberg-Marquardt scheme to obtain the *x*, *y* and *z* localization.

2.1.2. Multi Emitter Based Fitting Algorithms

MLE fitting for multi emitter

A MLE based method was developed by Huang *et al* [26] which also used a Gaussian model fitting to perform multiple emitter fitting in a given sub-region containing many emitters. First filtering is done to identify sub-regions with cluster of emitters in them and then the positions of *N* emitters are found sequentially where the *N* emitter model uses information from the *N-1* emitter model. For the first emitter localization in a sub-region an initial estimate is made from the center of mass and for subsequent emitters the position information from the *N-1* model is used as initial position estimate. The remaining initial position estimate is found out by calculating the residuum image generated after subtraction of *N-1* model from data in the sub-region. If maximum intensity in the residuum is lower than a threshold then it is assumed that there are no more emitters left in the sub-region. Figure 2.1 shows how GPUGaussMLE algorithm works.

PALMER

PALMER (Parallel Localization of Multiple Emitters via Bayesian Information Criterion Recommendation) [68] is a fitting based multi-emitter localization algorithm which is based on the combination of GPU, parallel computation and model recommendation via Bayesian Information Criterion which has

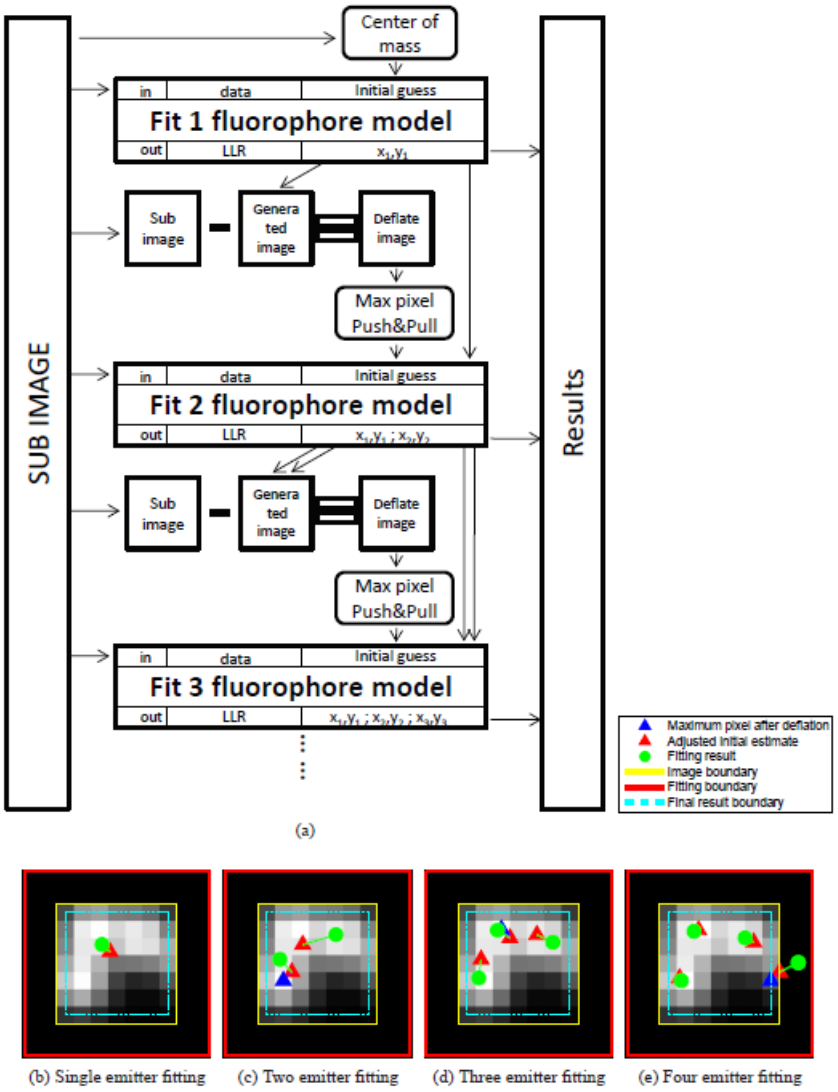


Figure 2.1: [GPUGaussMLE [26]] This figure gives an over-all schematic representation of the working of the GPUGaussMLE algorithm.

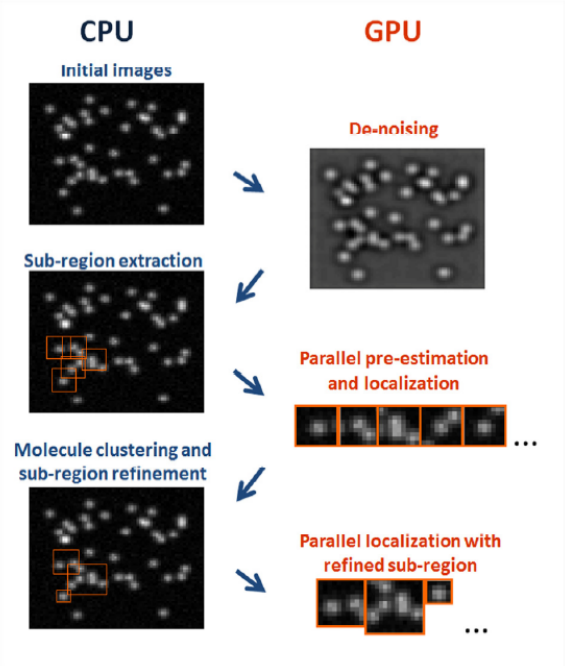


Figure 2.2: [PALMER [68]] Load sharing by CPU and GPU in PALMER algorithm.

a very low false positive rate. It is robust giving very precise localization over a wide range of emitter densities. This algorithm however cannot be used to perform 3D localization.The PALMER method is based on the former method SSM_BIC [55].The assumption is that each emitter is independently contributing to the observed signal at each pixel, the imaging model for multiple emitters is the convolution of PSF and function with respect to the positions and fluorescence intensities. The positions of emitters can be obtained using MLE and gradient descent at a sub-pixel level.

The algorithm pre-estimates the number of emitters and their position by finding the local maxima in the sub-region and if the centre and 4 connected neighbours are higher than the threshold the centre pixel is added to the list of potential candidate emitters and the process is repeated till there are no emitters left in the sub-region or the maximum number of emitters defined by the user has been reached.

CPU and GPU both are used in this localization procedure. The initial images are loaded into the computer's memory and convolution is performed with an averaging and annular filter to de-noise the images. Then sub-regions are cut-out from the image with a size of 9x9 using 5 times the background value for the threshold to localize maxima and the sub-region was removed from the original de-noised image to ensure that other sub-region cutting could be done independently. The algorithm then pre-estimates the number of emitters and their position by finding the local maxima in the sub-region and if the centre and 4 connected neighbours are higher than the threshold the centre pixel is added to the list of potential candidate emitters and the process is repeated till there are no emitters left in the sub-region or the maximum number of emitters defined by the user has been reached. A series of the model is generated based on the pre-estimate data and the optimum model is selected using Bayesian Information Criterion statistics. This is done on a GPU and the selected model is loaded into the CPU and then the mean-shift algorithm [71] is used to update the global position of the emitter after each iteration. Another round of emitter fitting is performed on GPU to improve sub-pixel localization and model recommendation. Figure 2.2 shows how the workload is shared between CPU and GPU

2D and 3D DAOSTORM

Holden *et al* [23] came up with a method which drew inspiration from an algorithm DAOPHOT 2 [64] which was used in astronomy. The new method called DAOSTORM can be used to localize emitters even when there is overlap amongst the PSFs. Initially candidate emitters are selected by using DAOFIND [18], an algorithm which convolves a truncated circularly symmetric Gaussian with the input image and local maxima in the image are selected as candidate emitters. Theoretically, the optimum

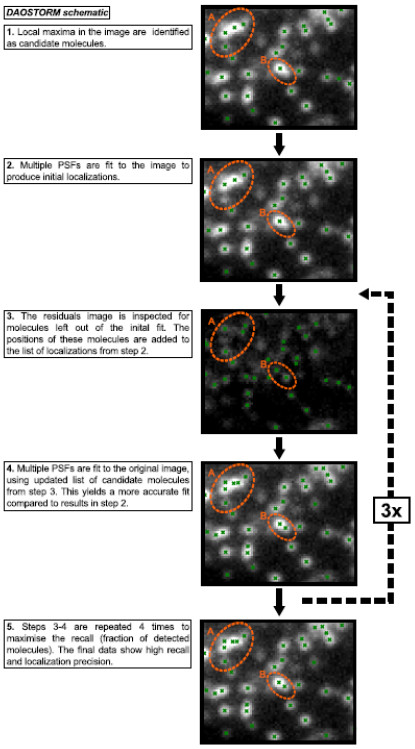


Figure 2.3: [DAOSTORM [5]] Flow of the DAOSTORM algorithm

way of localizing multiple molecules in a high density frame is by global minimization of a fit of all model PSFs in the whole frame but this is computationally very expensive as the time complexity is $O(N^3)$. If the whole image is split into approximately v number of sub-groups of non-overlapping cluster and simultaneous fitting is done for each cluster then the time complexity scales down to $O(N^2)$. The PSF model is generated keeping the size of the PSF constant so the fitting parameters are only the location and intensity which again makes the computation inexpensive. First the algorithm automatically groups overlapping molecule and fits each group with multiple model PSFs simultaneously. After initial fitting, a residual image is calculated containing the PSFs which were not identified during the initial localization step. Peak finding is carried out in the residual image and the newly identified emitters are added to the list of emitters from the initial steps and the whole fitting of the PSFs is carried out again and the whole process is done until there are no more unidentified emitters in the residual image. Both the method claim to have a very high recall rate even when the particle density is very high and across a broad range of SNR but both the algorithms are computationally very expensive.

Babcock, Sigal and Zhuang [5] modified the DAOSTORM algorithm to make it compatible for 3D localization. The primary difference between 2D DAOSTORM and 3D DAOSTORM are :-

- 2D DAOSTORM fits the image with a fixed shaped PSF while 3D DAOSTORM fits the images with elliptical Gaussian with varying x and y width which is dependent of the axial position.
- The error in fit is calculated using maximum likelihood estimator suitable for a Poisson distribution of error instead of Gaussian distribution of error.
- 2D DAOSTORM groups cluster of overlapping PSFs together and fits them simultaneously. In 3D DAOSTORM a single cycle of fit image based on updated position of every emitter.
- 2D DAOSTORM involves a cubic spline to correct for deviations in PSF from idealized Gaussian but 3D DAOSTORM doesn't.

Figure 2.3 shows how multi-emitter fitting is performed by DAOSTORM

2.1.3. Non-Iterative Methods

QuickPALM

QuickPALM [22], a free open source software available as a plugin for ImageJ that uses a modified centre of mass-based technique to localize single emitters and all single emitter utilizes a single thread and the parallel computation helps in reducing the execution time of the algorithm. QuickPALM's biggest advantage is it's speed and simplicity while it is not the most accurate method to perform localization as it is prone to errors when noise and background is high. Within each thread, a new unprocessed image is opened and the noise level of the unprocessed image is estimated by calculating the standard deviation from a 13x13 region centred on the minimum intensity pixel. This minimizes the chance of a single emitter pixel being present in the noise estimation operation. After this process, a bandpass filter similar to the ROI extraction algorithm is used to suppress noise and correct for the background.

The image is then searched for the maximum intensity pixel. Once the maximum intensity pixel is located a window is created using the maximum intensity pixel as the centre and the length of the window is twice the FWHM (Full Width at Half Maximum) defined by the user. The local SNR (Signal to Noise Ratio) is calculated by dividing the mean intensity within the window by the relative noise standard deviation calculated beforehand assuming the noise levels are same across the image. If the SNR calculated is less than the pre-defined SNR then the process is halted else the candidate spot is run through a series of tests passing which the spot will be registered as a single emitter. The tests are:-

- If the spot is not a part of the image edge
- If the intensity is not saturated
- If the spot doesn't overlap with any previous spot

If the tests are passed then the centre of mass of the spot is calculated using equations :

$$c_x = \frac{\sum_{i,j} s_{i,j} x_{i,j}}{\sum_{i,j} s_{i,j}} \tag{2.9}$$

$$c_y = \frac{\sum_{i,j} s_{i,j} y_{i,j}}{\sum_{i,j} s_{i,j}} \tag{2.10}$$

where c_x and c_y are the co-ordinates of the centre of the mass, $s_{i,j}$ is the intensity of the pixel and $x_{i,j}$ and $y_{i,j}$ are the co-ordinates of the pixels. The spot shape is then calculated using the following parameters stated in the following equations :

$$\sigma_l = \frac{\sum_{i,j} (c_x - x_{i,j})}{\sum_{i,j} s_{i,j}} \tag{2.11}$$

$$\sigma_r = \frac{\sum_{i,j} (x_{i,j} - c_x)}{\sum_{i,j} s_{i,j}} \tag{2.12}$$

$$\sigma_a = \frac{\sum_{i,j} (c_y - y_{i,j})}{\sum_{i,j} s_{i,j}} \tag{2.13}$$

$$\sigma_b = \frac{\sum_{i,j} (y_{i,j} - c_y)}{\sum_{i,j} s_{i,j}} \tag{2.14}$$

Where the sum of σ_a and σ_b gives the height of the PSF and the sum of σ_l and σ_r gives the width of the 2D spot. Once the length of the quadrants are known the height and width of the bead can be computed using equation 2.15 and 2.16

$$FWHM_x = \frac{2.354(\sigma_l + \sigma_r)}{2} \tag{2.15}$$

$$FWHM_y = \frac{2.354(\sigma_a + \sigma_b)}{2} \tag{2.16}$$

Figure 2.4 shows the workflow of QuickPALM algorithm.

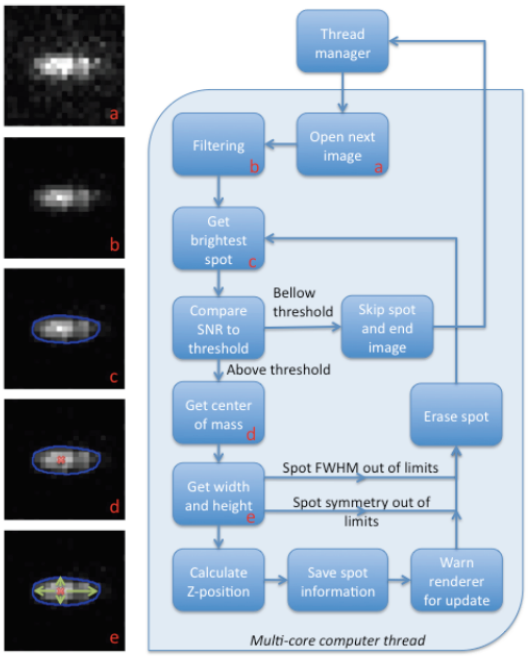


Figure 2.4: [QuickPALM [22]] Workflow of QuickPALM algorithm.

Radial Symmetry

Parthasarathy [53] presented a non-iterative, non-fitting based algorithm which can be used to localize the centre of a PSF utilizing the assumed radial symmetry of the PSF and can be used to perform 2D localization of PSFs. It is a computationally light, fast method which doesn't need a GPU and can reach performance of fitting based method. It also has a lot of drawbacks as the assumption of radial symmetry doesn't hold when PSF engineering is used to break symmetry and hence it is not suitable for 3D localization. It is also prone to the presence of trails of other PSF in the ROI and hence the dataset should be very sparse which is not always possible in real life setting. The idea behind the method is that for a radially symmetric PSF, any line which is drawn parallel to the gradient of any point will intersect at a point which the centre of the PSF as shown in Figure 2.5. The distance between any such line and the centre of the PSF will be theoretically zero. In an image which has limit pixel size and noise present in the image, the centre is the point which minimizes the distance of the centre to all such lines. In this method, the Robert cross operator is used to compute the gradient of the image. The advantage of using the Robert cross operator [14] is that it helps to compute both the components of the gradient at the same point simultaneously instead of computing the *x* and *y* component separately and the slope of the gradient can be computed using :

$$m_k = \frac{(I_{i+1,j+1} - I_{i,j}) + (I_{i,j+1} + I_{i+1,j})}{(I_{i+1,j+1} + I_{i,j}) - (I_{i,j+1} + I_{i+1,j})} \tag{2.17}$$

and any of the grid midpoint located at (x_k, y_k) a line with the slope m_k passing through the grid midpoint can be written as :

$$y = y_k + m_k(x - x_k) \tag{2.18}$$

The distance between a point and a given line is described using simple geometry and the problem is defined to find the best fit point (x_c, y_c) by minimising the function stated by :

$$\chi^2 \equiv \sum_k d_k^2 w_k \tag{2.19}$$

where d_k is the distance between a line and a particular point (i.e, the best-fit centre point) and the w_k is a weighting factor. This equation can be solved analytically and the output is the co-ordinate of the PSF centre.

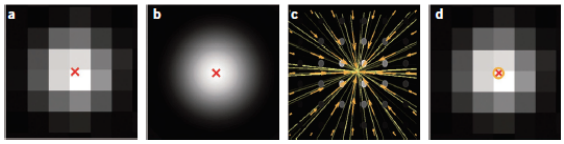


Figure 2.5: [Radial Fitting [53]] The concept of radial fitting.

Gradient Fitting

Ma *et al* [39] proposed a gradient-based method which could be used to localize PSFs in 3D Astigmatism microscopy in a non-iterative method. It is very fast computationally and can be used for real time particle tracking but MLE based fitting methods are 3-4 times more accurate than this method. The imaging model used is defined by:

$$I(m,n) = \frac{N}{2w_x w_y}\exp[-(\frac{(m-x_c)^2}{2w_x^2} + \frac{(n-y_c)^2}{2w_y^2})] \tag{2.20}$$

where (x_c, y_c) is the lateral centre position of the bead and (w_x, w_y) is the width of the bead in x and y direction and (m,n) are the co-ordinates of the lateral plane. The exact gradient along x and y axis can be computed using the partial derivatives of equation 2.20 which can be estimated using the following equation

$$G_x = I(m,n)\frac{-(m-x_c)}{w_x^2} \tag{2.21}$$

$$G_y = I(m,n)\frac{-(n-y_c)}{w_y^2} \tag{2.22}$$

The limitation in real time application is that the images are not free of shot noise and have limited pixel size. This makes the computation of the exact value of G_x and G_y difficult using equation 2.21 and 2.22. To overcome this, 2 optimized gradient operators are used to compute g_x and g_y by convolving the kernels with the raw image A. The measured gradient matrix g gives a good approximation for the theoretical gradient G given by :

$$g_x = \begin{bmatrix} -1 & 0 & 0 & 1 \\ -2 & 0 & 0 & 2 \\ -2 & 0 & 0 & 2 \\ -1 & 0 & 0 & 1 \end{bmatrix} * A \tag{2.23}$$

$$g_y = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & -2 & -2 & 1 \end{bmatrix} * A \tag{2.24}$$

The measured gradient may still deviate from the theoretical gradient G which is shown in Figure 2.6 therefore the non-linear least square method is used to calculate the best fit G with the least deviation D to the measured g. The metric deviation D is calculated as an angle θ which is defined by :

$$\theta \approx \sin\theta = \frac{|G_y.g_x - G_x.g_y|}{|G||g|} = \frac{|e(x_c-m)g_y - e(y_c-n)g_x|}{\sqrt{e^2(x_c-n)^2 + (y_c-m)^2}.\sqrt{g_x^2+g_y^2}} \tag{2.25}$$

where ellipiticity e is defined as $(w_x/w_y)^2$. The total deviation D can be computed using :

$$D = \sum_{m,n}\theta^2.W \approx \sum_{m,n}\frac{(e(x_c-m)g_y - e(y_c-n)g_x)^2}{(e_0^2(x_0-n)^2 + (y_0-m)^2)(g_x^2+g_y^2)}.W \tag{2.26}$$

where e_0 and (x_0, y_0) are initial estimates obtained using the centroid method and W is a weighting fraction. Mathematically, D is minimum where the partial derivative is equal to zero and x,y and e

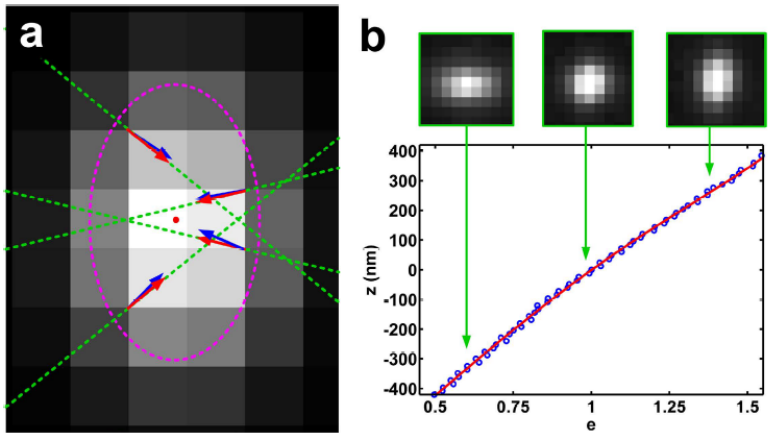


Figure 2.6: [Gradient Fitting [39]] a). This figure show how the actual gradient (green) deviate from the computed gradient (red and blue). b). Mapping of eccentricity to axial position.

can be obtained by solving closed form equations. The algorithm outputs an ellipticity value which is mapped to an axial position using a calibration curve which is generated using an iterative fitting algorithm which is more accurate.

Phasor based method

Martens *et al* [40] proposed a method which performs localization with precision comparable to the most accurate fitting based methods and it is computationally inexpensive and can be run on CPU. The method involves calculating the first Fourier co-efficients in *x* and *y* directions to create a vector which represents phasors. The angle of the phasor is used to identify the centre of the PSF and ratio of the magnitude of the x and y component is used to compute the z position of the PSF and similar to the gradient fitting method [39] the ratio of magnitude maps to a z position using a calibration curve.

FluoroBancroft

Anderson *et al* [2] came up with a localization algorithm which draws inspiration from the Global Positioning System where three measurement positions are used to determine the 2D position of the user using a technique called triangulation shown in Figure 2.7. It is an algorithm which is very fast but this algorithm works on very sparse dataset and 3D localization is not possible with this algorithm. The emitter localization problem is also designed in a similar manner, where the true location of the emitter can be found from the overlap of measurements. Similar to all the other direct methods it is a very quick method but the accuracy isn't close to the state of the art fitting methods. Here, the assumption is that the fluorescence intensity depends on the distance between measurement points (centre of the pixel in CCD array) and the position of the emitter. A single measurement results in a range which can be traced in the form of a circle. A measurement would result in another circle and the location of the emitter will be in the intersecting area of the two circles. Another measurement would yield in yet another circle of possible range and in absence of noise, the 3 circles would intersect at a point giving the exact location of the emitter's position. The presence of noise leads to error in the range predicted by each circle and all the three circles would not intersect at a single point and rather than an analytical solution an estimation has to be found. The algorithm models the emitter particle as a Gaussian profile with shot noise and Poisson noise. The half-value point of intensity is defined by the Rayleigh's radius. Equation 2.27 defines the emitter model

$$I(x,y) = me^{-\frac{r^2}{2\sigma_x^2}} + \eta_B + \eta_{sh} \tag{2.27}$$

where m is a scaling factor determined by the photon emission rate of the fluorophore, σ_x and σ_y are the width and height of the PSF and η_B being the background and η_{sh} is the shot noise and range from measurement points (x,y) to true position (x_o, y_o) is defined :-

$$r = \sqrt{(x - x_o)^2 + (y - y_o)^2} \tag{2.28}$$

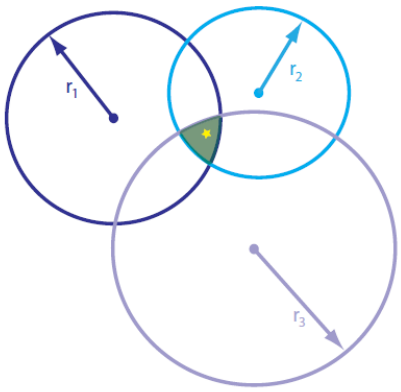


Figure 2.7: [FluoroBancroft [2]] Localization using the idea of 'triangulation'.

rearranging which we get :-

$$r^2 = 2\sigma_x^2 ln(m) - 2\sigma_x^2 ln(\langle I \rangle - N_B) \tag{2.29}$$

where r^2 gives the range of the measurement and $\langle I \rangle$ being the expected photon intensity and N_B being the expected background noise level.

Each measurement point results in a range equation which can be arranged into a system of an over-determined linear equation which is solved using Moore-Penrose pseudoinverse to obtain the position of the emitter.

Joint Distribution

Larkin and Cook [32] presented a method which uses the probability distribution of each photon to localize the emitter position by building up a joint probability distribution shown in Figure 2.8. These authors claim that this algorithm performs better than MLE when the SNR is very low but this algorithm isn't precise compared to MLE based fitting algorithm when the SNR is high and the algorithm has a fatal flow of assuming the brightest pixel as a candidate pixel. This is a non-iterative algorithm which is used to localize emitters which are imaged by a CCD. The assumption behind this work is that every photon registered on the CCD sensor carries information about the position of the emitter. This information is blurred by the PSF of the optical system. Here PSF is defined using many probability distributions - one for each photon registered on the camera. Assuming all the photons coming from the same single emitter a joint probability distribution is built up and used to localize the emitter's position. This algorithm provides a closed-form solution and is robust against noise. Generally, photons falling on the camera pixel are described as a binning action like building a histogram where the sub-pixel spatial location is lost. In conventional microscopy when a photon is registered in a pixel the probability distribution of the photon is defined by a step function where inside the bounds of the pixel the probability is 1 and outside the pixel, the probability is 0. In this approach, the probability distribution is treated as a Gaussian distribution where the probability is spread into neighbouring pixels as well. Each photon has an individual Gaussian probability distribution. All the distribution are used a joint probability distribution and the mean of the joint probability distribution is the location of the centre of the emitter and the variance of the normal joint probability distribution is the width of the PSF. The closed-form equation 2.30 describes the mean of the joint probability distribution which is the estimate of the emitter's true position

$$\mu_o = \sum_{i=1}^{N} (\mu_i \sigma_i^{-2}).(\sum_{i=1}^{N} \sigma_i)^{-1} \tag{2.30}$$

3D-PALM

York *et al* [72] described a way to estimate emitter position from images based on cross correlation without using a PSF model or utilizing the knowledge of the optical configuration. One of the biggest

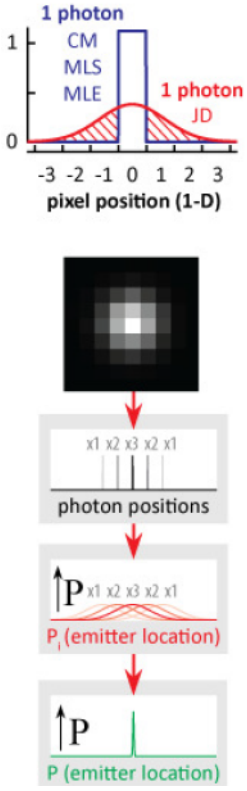


Figure 2.8: [Joint Distribution [32]] The concept behind the joint distribution algorithm for localization.

advantages of this method is that no prior assumption is needed about the PSF generation model but the biggest drawback is it's complexity. The various steps of the algorithm are listed below :

- Construction of a calibration stack
- Candidate particle selection
- Localization
- Drift Correction
- Link Localization and re-localization (optional)
- Construction of the image histogram

To construct the calibration stack the image stack is cut out in such a way that it contains only one fiducial marker with minimal blank space surrounding the marker. All images at the same piezo-position is averaged out and the image is smoothened with a Gaussian filter.

After the calibration stack is constructed the image is filtered with a Gaussian-Laplace filtered to remove slowly varying background and quick-varying noise. All the pixels above a user defined threshold are marked and a rectangle the size of the calibration stack xy dimension is placed around the brightest spot and the rectangle's location is recorded and all the pixels inside the rectangle are unmarked and this process is done for the entire image recording the position of all the candidate pixels.

The previous image is subtracted from the current frame to generate a differential image which helps to identify the candidate emitter's 'birth' where brightness increases significantly from the previous frame and correspondingly the candidate emitter's 'death' is identified where the brightness decreases in the current frame compared to the present frame.

Cross-correlation is then used to generate the similarity index between the candidate image and the calibration image after cross-correlation and the candidate image is then shifted in x, y and z to find

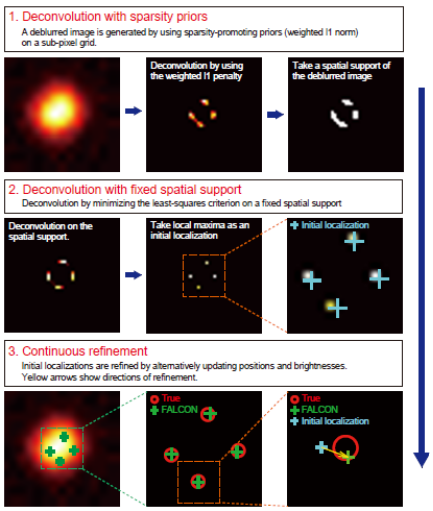


Figure 2.9: [FALCON [42]] The 3 stages of working of the deconvolution algorithm FALCON

the position where the similarity index is highest and the candidate image is position is identified by adding the same amount of shift by which the calibration image was shifted and interpolation is used to shift the image on a sub-pixel scale and the final 3D co-ordinate is recorded.

FALCON

FALCON [42] is a deconvolution algorithm which performs localizations of multiple emitters present in a sub-region. It is a complex algorithm which is computationally very expensive but this algorithm is a robust multi-fitting algorithm which performs very well over a broad range of physical conditions. In deconvolution based localization, a molecules position is estimated by upsampling and subsequent deblurring of a low resolution image. In this method, a deblurred image is generated by using a sparsity promoting priors on a sub-pixel grid. This is followed by minimizing a least square criterion on a sub-pixel grid which is followed by a finer refinement process to obtain the final localization list of multiple emitters in a sub-region of a low resolution image. Figure 2.9 pictorially describes idea behind the FALCON algorithm.

Wedged Template Matching

Takeshima *et al* [65] came up with a method to increase the temporal resolution in localization microscopy by introducing a template-based multi-emitter fitting algorithm. The algorithm is capable of working with frames which have high emitter density where PSFs are overlapping. The method is called Wedged Template Matching where the algorithm uses a template for a single molecule to make the best estimates. Partial image templates are used to recognize the single emitter and make coarse localization and the full template is used to remove the single emitter from the cluster of overlapping PSFs. WTM is a deconvolution method where the idea is to match a segment of the model function of the PSF to the template and even in the region of severe overlap the model matches the template accurately near the edges and this helps in identification of the molecule as a candidate emitter. WTM algorithm comprises of 4 stages:-

- Preparation of the single-molecule model
- Background subtraction
- Wedged template matching: Camera level and Sub-pixel level
- Emitter removal from the cluster

For the preparation of the model, a Gaussian model is used with a diameter of PSF and total intensity of the PSFs as parameters. After defining the model the frame is convolved with a convolution mask of approximately 5x5 pixels and then the pixels which have a value higher than the defined intensity

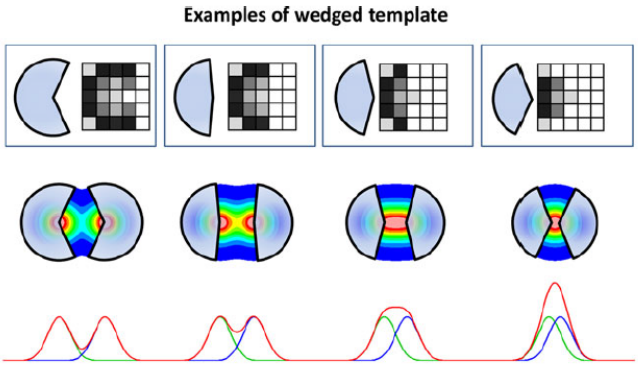


Figure 2.10: [Wedged Template Matching [65]] Example of different templates used for template matching.

threshold are considered as candidate emitters. The WTM algorithm prepares 4 different templates to deal with various degrees of PSF overlap. The template is then placed at the location which is identified as a candidate emitter and normalised cross-correlation coefficient is computed and then the template is rotated again in step sizes of 45 and the correlation coefficient is computed again. The pixel with the highest similarity index (for any angle θ) is selected as the emitter pixels. This process is done at camera resolution level and the higher precision template matching is applied to the emitter pixel on a sub-pixel resolution grid. Before applying the finer matching, the desired sub-pixel resolution of either (9x9 or 15x15) is selected within a single pixel. For each camera pixel, the same number of templates are created and then a similar matching process is done at all the sub-pixel grid points and the highest similarity index sub-pixel is identified as the true centre. Once the true centre is located on the sub-pixel grid the single emitter is subtracted from the cluster using the PSF model and the intensity of pixel which were are taken as parameters while building the model for template matching and the process is repeated until there is no candidate pixel that remains in the residuum image. WTM then finally outputs a list of localized true centres. Figure 2.10 shows the example of the different templates which are used for the wedged template matching algorithm. The biggest advantage of using this algorithm is that this is designed to handle cases of severe overlap and tends to give results comparable to state of the art multi-emitter fitting algorithm and it's biggest drawback is that no templates are available to perform 3D localization.

2.2. Deep Learning and Localization Microscopy

In this section, the emergence of deep learning in the domain of localization microscopy is discussed. This section starts off by presenting deep learning based methods which could identify and count single molecules followed by discussion on the development of various methods which started addressing different sections of the localization microscopy pipeline which includes background estimation, localization and super-resolution image reconstruction.

Identification of Single Molecule

The emergence of deep learning in the single-molecule localization microscopy is relatively new when compared to the conventional localization schemes. One of the earliest works, in single-molecule localization microscopy, is by Powen *et al* [9], where an artificial neural network-based approach has been developed to identify the single molecule. In this paper, the identification of single molecules has been designed as a classification problem. The advantage of using a neural network over a pre-defined model is that if the model selection is not done accurately it can lead to results which distort the reality even though neural networks don't provide insight into the underlying mechanism. The neural network architecture is designed to tackle a two-class problem and can identify two different fluorophores. For the purpose of experiments 3 different type of fluorophores [17] [1] [73] was used. The neural network used in this paper is a simple two-layer feed-forward network with an activation function being a log-sigmoidal function which a popular choice of a transfer function in a classification problem. The neural

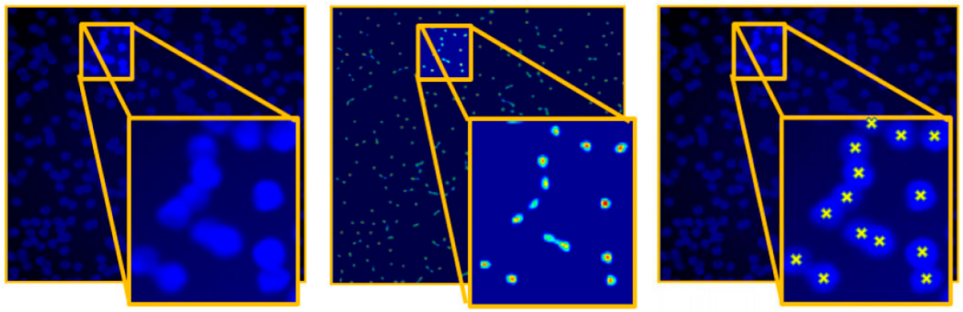


Figure 2.11: [Cell Counting and Detection [70]] The image on the left shows the low resolution image containing PSF cluster. The center figure shows a density map generated by the network and the network on right shows cell counting using the maxima of the density map.

network has 64 input nodes and two output node with output values of 0 and 1 at each node, [0,1] denoting the identified fluorophore is of the type-1 and [1,0] suggesting that the fluorophore is of type-2. This paper is one of the earliest works which show that a neural network based approach can be used to tackle various sub-tasks of SMLM.

Cell Detection and Counting

Xie, Noble and Zissermann [70] came up with a method for automated cell-counting and detection in microscopy images. In this work, a fully convolutional network is set up to regress a cell spatial density map as an output. This is an alternative approach which can be used to detect and count cells when traditional segmentation algorithms can't be applied due to cell clumping and overlap. The cell counting problem can be solved by two approaches:

- Detection based counting. In this method, prior detection or segmentation is required
- Density estimation. No prior segmentation is required.

The CNN based method of Xie,Noble and Zissersmann doesn't require any prior segmentation and approaches the cell counting problem using a density estimation based method and shows that cell detection can be used as a side benefit of the cell counting task. The problem is mapped as a supervised learning task which maps an image *I(x)* to a density map *D(x)* for a *mxn* image. The CNNs are trained on synthetic data and their performance is evaluated using experimentally collected data. To train the neural network, the ground truth is presented in the form of a dot annotation and each cell is represented as Gaussians and the density map is formed by the superposition of the Gaussians. The network regresses a density map and then local maxima are counted to count the number of cells present in the image frame. Since the cells are much smaller than the image frame the need of networks which can represent highly semantic structure is not needed and a simple CNN network is sufficient. Inspired by very deep VGG-net [58] small kernels of size 3x3 or 5x5 is used. The number of feature maps in a higher layer is increased to counter the loss of spatial information by pooling. To increase training data size, the synthetic images are cut into smaller frames of 100x100 pixels and simple data augmentation techniques such as flipping and small rotations are also employed and the images are normalized. The cost function used by the network is defined by :-

$$I(W; X_0) = \frac{1}{M} \sum_{i=1}^{M} (Y_n - X_n^i)^T (Y_n - X_n^i) \tag{2.31}$$

where W are all trainable parameters, X_0 is the input patch and Y is the ground truth with annotations and M is the total number of training data. Stochastic gradient descent with momentum is the choice of optimizer for this problem. Figure 2.11 shows the input and output of the network. The generation of density map with maxima information can be utilized to segment an emitter before performing localization.

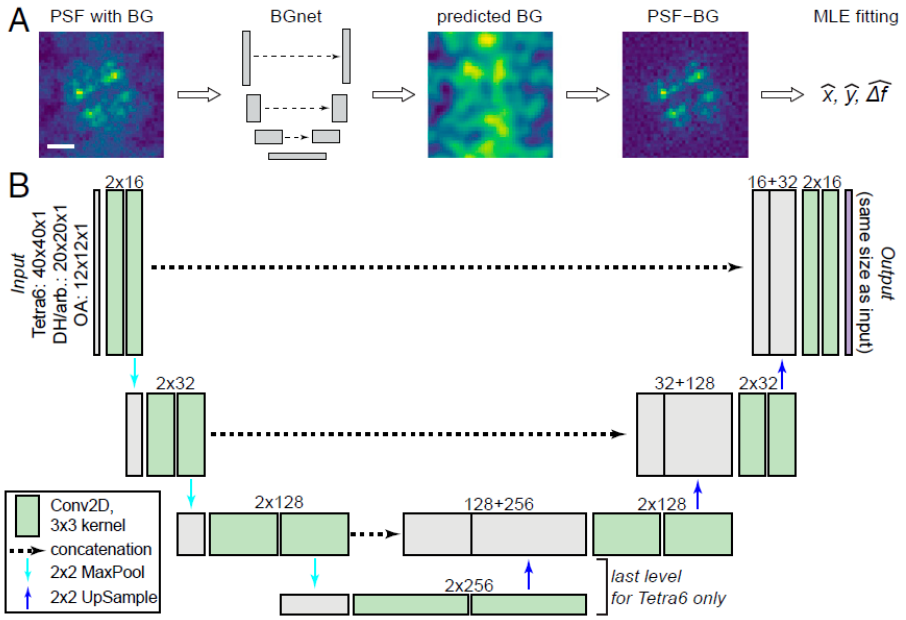


Figure 2.12: [BGnet [44]] a). This image shows the input to the network and the estimated structured background. b). This depicted the network architecture of BGnet.

Background Estimation and Data Augmentation

Mockl *et al* [44] came up with a deep neural network called BGnet which can estimate background from optical images. In single-molecule localization microscopy, the quality of the rendered images is dependent on the accuracy of localization of the single molecules and the presence of background can hamper the quality of localization. The fitting algorithm can handle the presence of constant background which is treated as a constant offset but the presence of structured background is much more detrimental to the localization process. Treating a structured background as offset makes localization inaccurate as the underlying structures alter the PSFs. Any approach to remove the presence of structured background involves the identification of the structure and the ability of the algorithm to differentiate between the PSF and the structured background.

BGnet allows fast and accurate estimation of the structured background. The network is heavily inspired by U-net [57], a U shaped convolutional encoder-decoder network which is very popular in biomedical segmentation applications. The reason such an architecture is chosen is that the segmentation task is very similar to the structured background estimation task where a structure is embedded in a structured background and the underlying structure has to be removed. The idea is to first reduce the spatial size of the image while increasing the filter space and after condensation increasing. For training the network, a dataset is provided which contains PSFs at the various axial position and different frequencies of structured background where the PSFs are simulated using vector diffraction theory. Figure 2.12 shows the input and output of the network and the network architecture.

Another use of U-net has been described by Schmidt *et al* [69] to generate training data for image restoration in optical microscopy using deep learning.

Localization using Deep Learning

Zelger *et al* [74] use a deep convolutional network based on the VGG-16 network to localize single molecules in 3D. The network has one or two convolution layer followed by a pooling layer to extract the most important spatial information from the previous layers. The kernel size is fixed at 2x2 pixels because the objects of interests are not very large and the number of kernels has been increased to 1024 to extract the maximum amount of spatial information from the PSFs. The network uses the ReLu activation function which is suitable for a regression problem and the Adam optimizer to update the weights by back-propagation. The training data comprises 10,000 images of emitters at random 3D position varying between -1.3 μm to 0.1 μm (arbitrary choice made by authors) in the z dimension

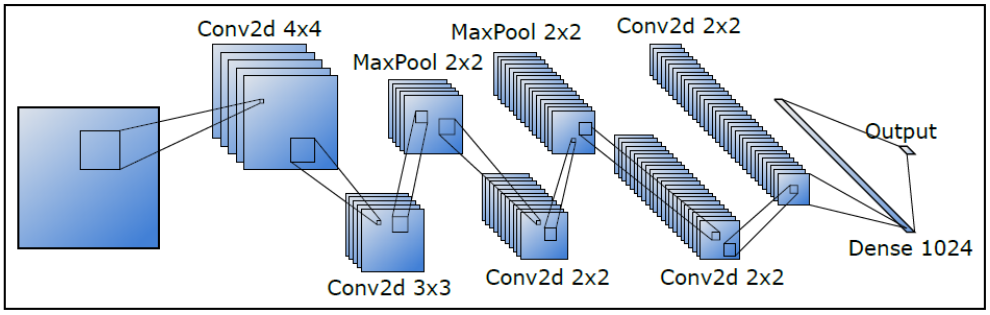


Figure 2.13: [Localization [74]] Architecture of the proposed network.

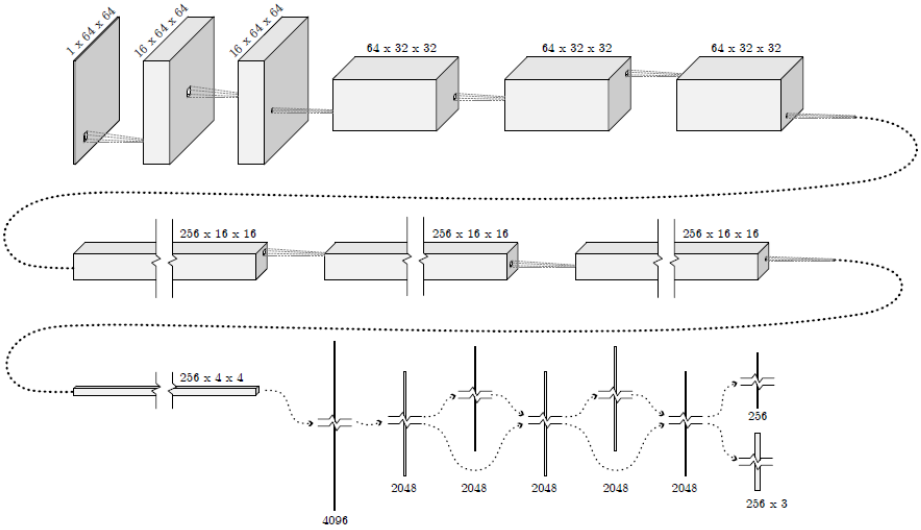


Figure 2.14: [DeepLoco [10]] Architecture of the DeepLoco network.

and x,y range of [-0.5μm,0.5μm]. The signal photon count was randomly changed from 1000 photon to 9000 photons and the background was varied between 75 to 250 photons. The training images are generated from ROIs containing single emitters extracted from a single frame by segmentation of the sparse images and the neural network is used to predict the sub-pixel position. Figure 2.13 shows the architecture of the proposed network.

Boyd *et al* came up with a method named DeepLoco [10] shown in Figure 2.14 which uses a deep neural network to directly map images with emitters to their respective locations. The performance of the network is then evaluated using a novel loss function developed by the authors. The advantage of this method is that this method can be applied to arbitrary aberrations, noise and non-linearity. The most important contribution of this work is the novel loss function which the authors came up with.

Aritake *et al* [4] describe a neural network (see Figure 2.15) approach to perform single-molecule localization in a multi-plane setup. The problem with multi-plane setup is that even the smallest drift is detrimental to the localization process and the authors formulate an approach to describe the 3D localization problem along with the estimation of lateral drift as a compressed sensing problem. In this work, the authors did not use PSF engineering to encode axial position but instead used a quad plane microscope and it's focal points to encode the position. In such setup, lateral drifts are very detrimental and the assumption is made that all estimation might have some level of sub-pixel drift in the image. To make the estimation accurate the estimation of the position along with the lateral drift estimation is important and this is formulated as a compressed sensing problem. The CNN employed in the paper is based on the FSRCN network [15] which accurately predicts the position of the molecule and is robust against drift and does not need any explicit drift correction. The imaging model can be defined :

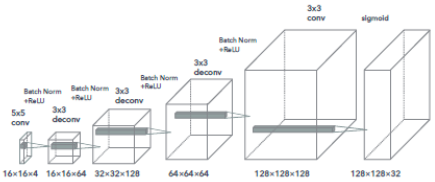


Figure 2.15: [Localization using Deep Learning [4]] Architecture of the proposed network to localize single emitter.

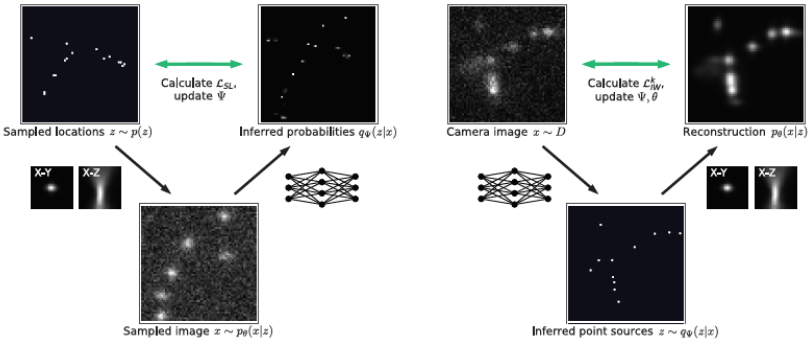


Figure 2.16: [Localization using Deep Learning [61]] This figure gives an over-all schematic representation of the experiment

$$y^l = Hw + \epsilon \tag{2.32}$$

where y^l is a vector of fluorescence intensity or the observed low-resolution image and H is the observation matrix containing the PSFs and w is the molecule distribution along with noise models. Now the problem is to estimate the elements of the w matrix from the low-resolution image y^l and H is an over-complete matrix. This can be re-written as the following equation which is called compressed sensing problem :

$$minimize_w ||y^l - Hw||_2^2 + \lambda ||w||_1 \tag{2.33}$$

Speiser, Turaga and Macke [61] proposed a method (DECODE) which uses a temporal context from multiple sequentially imaged frames to detect and localize molecules. It is a mixture of supervised and unsupervised learning which makes it robust against a mismatch in the generative model. It addresses the vulnerability of DeepSTORM [45] and DeepLoco [10] algorithms which uses so-called simulator learning where due to the want of training data the models are trained on simulated data and deployed on experimental data. Imperfection in the simulation model can induce imperfection in the localization process. The authors propose a method which combines simulator learning with variational auto-encoder to make the networks robust. The DECODE network learns hidden features from consecutive images and these frame specific features are integrated by a module which outputs five output maps. The first map is a binary map of particle detection, the second map predicts the brightness of detected particles and the final 3 maps outputs the spatial location of the particles.

DeepSTORM3D [46] is a neural network-based approach which is designed to localize overlapping PSFs over a large axial range and output a list of 3D positions. The second contribution made by the authors is the development of a PSF for 3D localization of dense emitters over a large axial range. Simulated emitter positions are fed to the imaging model with the new developed PSF which generates a low-resolution image which is then fed into the trained network which outputs the 3D positions of the localized emitters. The CNN structure has a multi-scale context aggregation module to process the input low-resolution images and it extracts features using a growing receptive field. It is followed by an upsampling stage and the last module refines the lateral and axial position of the emitters and generates the predicted vacancy grid shown in Figure 2.17.

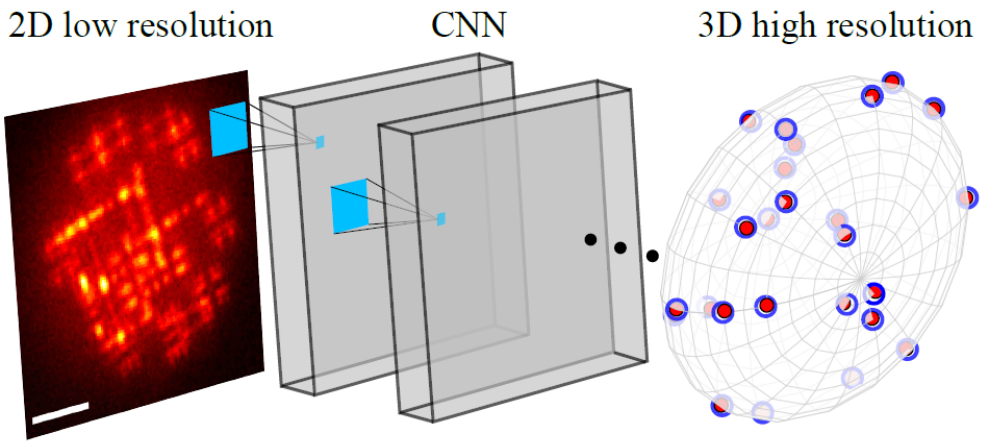


Figure 2.17: [DeepSTORM3D [46]] Representation of the working of DeepSTORM3D algorithm which generates the co-ordinates of the localized emitters.

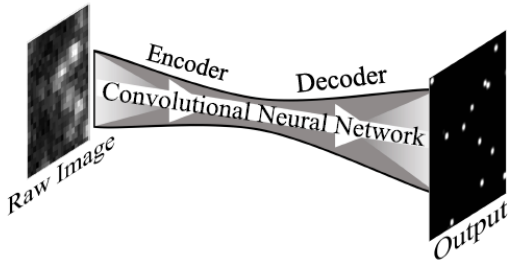


Figure 2.18: [DeepSTORM [45]] Representation of the working DeepSTORM algorithm which reconstructs super-resolution images directly from low resolution images.

Image Reconstruction

Nehme *et al* [45] presented a deep learning approach to directly take low-resolution images as input and give out a super-resolution image as output. The network is based on the fully conventional encoder and decoder architecture which is inspired by the cell counting network [70]. The network first aggregates spatial information at multiple scales in the encoding stage using multiple convolution stages and later in the decoding stage the spatial dimensions are restored using deconvolution stages. The final pixel-wise prediction is created using a depth reducing convolution filter with a linear activation function. Since the network doesn't produce a localization list and produces a direct super-resolution image shown in Figure 2.18, the loss function for training the net uses a regression approach. The squared l_2 distance between the network's prediction and the ground truth image which is formed by a set of delta spikes convolved with a 2D Gaussian. The training process also promotes sparsity by introducing a l_1 penalizer. The loss function for this method is given by :

$$l(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^{N} ||\hat{x}_i \circledast g - x_i \circledast g||_2^2 + ||\hat{x}_i||_1 \tag{2.34}$$

where $\hat{x_i}$ is the networks prediction, x_i is the ground truth, g is the Gaussian which is convolved with the ground truth and prediction and N is the number of sample in the training set.

Ouyang *et al* [52] demonstrated an artificial neural network-based reconstruction method named ANNA-PALM to reconstruct super-resolution views from sparse localization images or widefield images. This task was defined as an image restoration task and the challenge with such tasks are that an infinite number of solutions exists unless constraints are imposed to restrict the solution to a lower-dimensional manifold. Suitable manifolds exist because most images are redundant and can be approximated with a smaller number of degree of freedom than the number of pixels. ANNA-PALM makes use of this fact and restores an under-sampled image in the time domain into a high-resolution

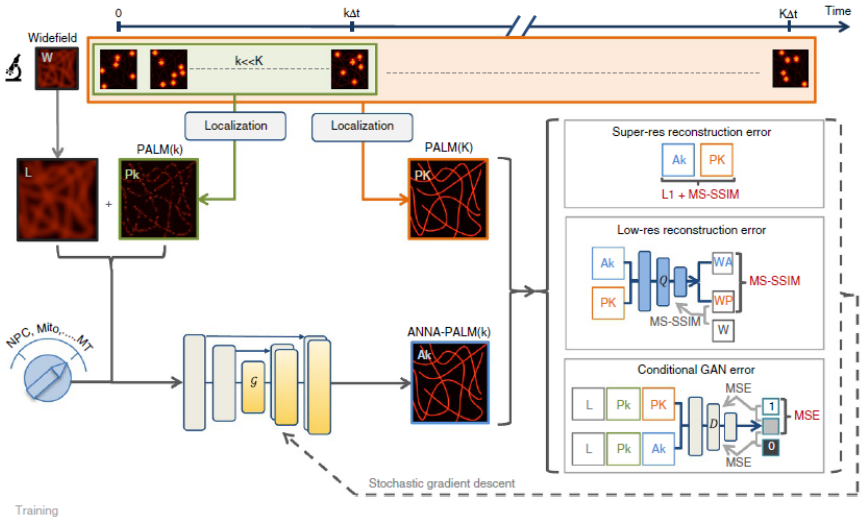


Figure 2.19: [ANNAPALM [52]] End to end representation of ANNAPALM algorithm for image representation which has a auto-encoder network and a Generative Adversarial Network.

image using an encoder-decoder network and a generative adversarial network. The CNN is similar to the U-net structure suitable for extracting multi-scale information and pixel-wise mapping and the GAN (generative adversarial network) network is used to process this information and generate the high-resolution image. Figure 2.19 shows the end to end representation of the ANNAPALM algorithm.

2.3. Unanswered Research Problems

The state of the art localization algorithms which do not use deep learning are mostly fitting based algorithms which are slow because of their iterative nature. The deep learning methods usually tackle one problem of localization microscopy as the parameter space scales up very fast if multiple problems have to be tackled. The deep learning method, smNet partially solves this problem by multiplexing each of the problem of aberration estimation, orientation estimation and 3D localization. The training process of smNet is done with an oversimplified model which is based on diffraction theory. In optical systems with a high NA objective lens, diffraction theory doesn't accurately represent the image formation process. In this thesis, the performance of smNet is characterized when smNet is trained with simulated data generated using a more accurate vector model. In this experiment, the concept of simulator learning is also tested which would remove the need for synthetically augmenting experimental data.

3

Physics of Image Formation

In this chapter, the image formation process and aberration in the imaging process are discussed. The concept of Cramer Rao Lower Bound which serves as a metric to compare the performance of any estimator is also discussed.

3.1. Diffraction Model - Scalar and Vector

Scalar Diffraction Theory - Wave Propagation

Diffraction theory is used to define the imaging process when the numerical aperture is small and paraxial approximation is possible ($sin\alpha$ = α). Diffraction theory was used by Zhang *et al* [76] to generate simulated images to train the neural networks. In diffraction theory, the polarization of light doesn't change after passing through a lens. Using Fourier optics, a light wave propagating from a point along the z-axis from $z = 0$ to $z = z$ can be represented using the following equation:

$$U(x, y, z) = \int\int df_x\, df_y\, \hat{U}(f_x, f_y, z) e^{2\pi j(f_x x + f_y y)} \tag{3.1}$$

where $U(x, y, z)$ represents the electric field distribution at a point $z = z$ and $\hat{U}(f_x, f_y, z)$ is defined as:

$$\hat{U}(f_x, f_y, z) = e^{\frac{2\pi jz}{\lambda}} e^{-\pi j\lambda z(f_x^2 + f_y^2)} \hat{U}(f_x, f_y, 0) \tag{3.2}$$

and $\hat{U}(f_x, f_y, z)$ represents the multiplication of the Fourier transform of the wave intensity at $z = 0$ with the propagation transfer function which is defined by the exponential terms in the equation. This formulation helps physicists approach the wave propagation problem from a systems perspective.

Scalar Diffraction Theory - Effect of Aperture

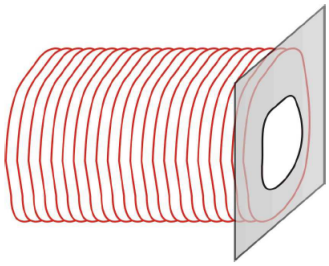


Figure 3.1: Representation of the interaction of a propagating light wave with an aperture [62].

When a propagating wave traveling from $z = 0$ to $z = z$ encounters an aperture as shown in figure 3.1 in the propagation path it gets diffracted by the aperture and the resulting electric field distribution at $z = z$ is represented using the equation:

$$U(x, y, z) = U_0 e^{\frac{2\pi j z}{\lambda}} \int\int df_x\, df_y\, \hat{T}(f_x, f_y) e^{-\pi j \lambda z (f_x^2 + f_y^2)} e^{2\pi j (f_x x + f_y y)} \tag{3.3}$$

where $\hat{T}(f_x, f_y)$ represents the Fourier transform of the aperture transmittance function.

Scalar Diffraction Theory - Effect of Lens

The presence of a lens in the propagation path modulates the phase of the incoming beam and changes the shape of the wavefront which is shown in figure 3.2. The equation 3.4 represents the modulation effect of the lens on the incoming wave where U_{in} and U_{out} represents the incoming and outgoing wave respectively and the exponential function is the lens modulation function. The equation 3.5 shows the effect of the lens as a transfer function multiplied to the Fourier transform of the incoming light wave.

$$U_{out}(x, y) = e^{-\frac{\pi j (x^2 + y^2)}{\lambda F}} U_{in}(x, y) \tag{3.4}$$

where λ is the wavelength of the light and F is the focal length of the lens.

$$U_{front}(x, y,) = \frac{e^{\frac{4\pi j F}{\lambda}}}{j\lambda F} \hat{U}_{back}(\frac{x}{\lambda F}, \frac{y}{\lambda F}) \tag{3.5}$$

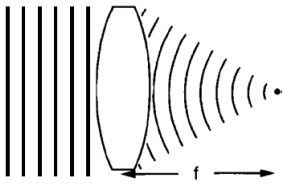


Figure 3.2: Effect of lens on the wavefront of a propagating light wave [62].

Scalar Diffraction Theory - Image Formation in a Microscope

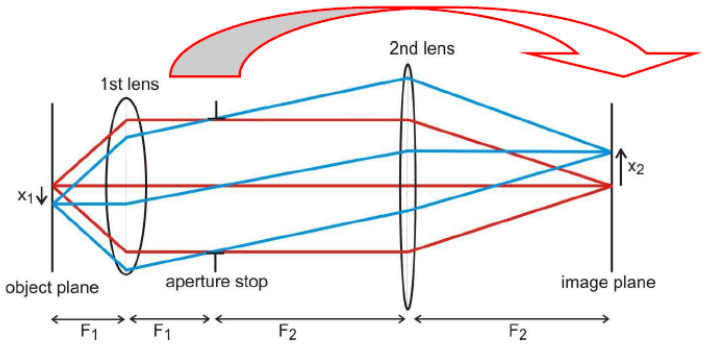


Figure 3.3: A telecentric 4f optical system [62].

Figure 3.3 shows a two-lens optical imaging system. The optical system is an example of a telecentric '4f' optical system where the aperture stop is imaged at ∞ in the image plane and numerical aperture (NA) and magnification (M) does not change with z position [62]. Using equations 3.1 and 3.4 the complex amplitude function at the image plane can be described using the following equation:

$$U_{image}(x_2, y_2) = -\frac{1}{\lambda^2 F_1 F_2} \int\int dx_1\, dy_1 h(x_2, y_2; x_1, y_1) U_{object}(x_1, y_1) \tag{3.6}$$

The term $h(x2, y2; x1, y1)$ is the point spread function (PSF) of the system. The detectors present in the image plane record image intensity which is found by taking the square of the complex amplitude. Since the emission from the fluorescent proteins is incoherent, image intensity at the detector is the convolution of the PSF with the fluorescence emission intensity of the object.

Vector Diffraction Theory

Figure 3.4 schematically represents the change in polarization when a light wave passes through a high NA lens. When working with a high NA lens the paraxial approximation ($sin\alpha = \alpha$) does not hold and the polarization of the incoming wave changes on interaction with a high NA lens [62]. The wave propagation from the back focal plane of the lens to the front focal plane of the lens is still represented by 2D Fourier transform but there are modifications made to the amplitude and polarization of the wave. The lens transfer matrix which is shown in equation 3.7 reflects how the input polarization changes when passing through a high NA lens. The equations 3.8-3.10 show the individual components of the propagating wave written as a Fourier transform incorporating the polarization change due to the high NA system [63][59].

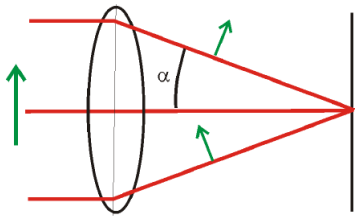


Figure 3.4: Representation of the change in polarization of light passing through a high NA lens.

$$\begin{bmatrix} E_x^{out} \\ E_y^{out} \\ E_z^{out} \end{bmatrix} = \begin{bmatrix} cos\theta cos^2\phi + sin^2\phi & (cos\theta - 1)sin\phi cos\phi & -sin\theta cos\phi \\ (cos\theta - 1)sin\phi cos\phi & cos^{\phi} + cos\theta sin^2\phi & -sin\theta cos\phi \\ sin\theta cos\phi & sin\theta sin\phi & cos\theta \end{bmatrix} \begin{bmatrix} E_x^{in} \\ E_y^{in} \\ E_z^{in} \end{bmatrix} \tag{3.7}$$

$$E_x(x', y') = \frac{1}{j\lambda F} \int\int dx\, dy \frac{cos\theta cos^2\phi + sin^2\phi}{\sqrt{cos\theta}} e^{-\frac{2\pi j(x'x + y'y)}{\lambda F}} \tag{3.8}$$

$$E_y(x', y') = \frac{1}{j\lambda F} \int\int dx\, dy \frac{(cos\theta - 1)sin\phi cos\phi}{\sqrt{cos\theta}} e^{-\frac{2\pi j(x'x + y'y)}{\lambda F}} \tag{3.9}$$

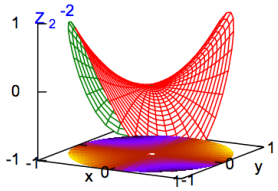
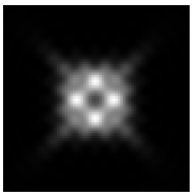
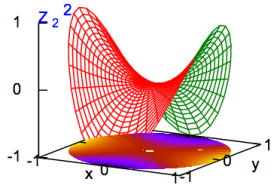
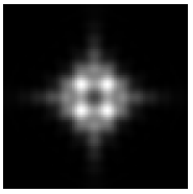
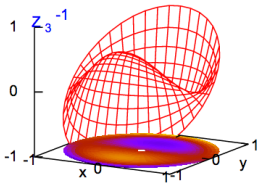
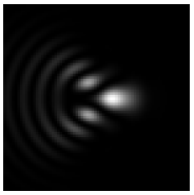
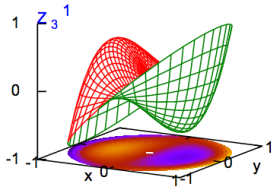
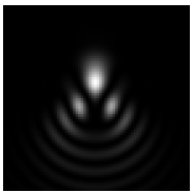
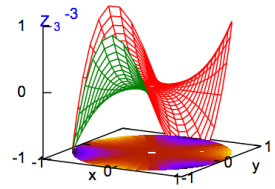
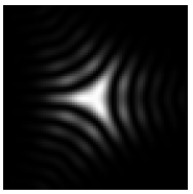
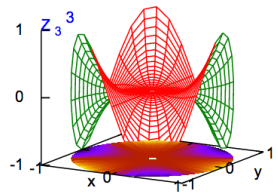
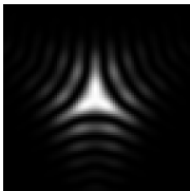
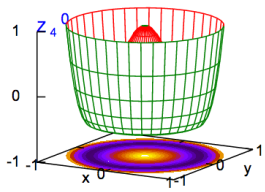
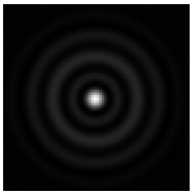
$$E_z(x', y') = \frac{1}{j\lambda F} \int\int dx\, dy \frac{sin\theta cos\phi}{\sqrt{cos\theta}} e^{-\frac{2\pi j(x'x + y'y)}{\lambda F}} \tag{3.10}$$

3.2. Aberrations and Zernike Polynomials

A perfect optical system produces a spherical converging wavefront which produces a sharp image at the sensor of the camera. Aberrations present in the optical system distort the spherical wavefront and results in a blurry image as seen in Figure 3.5. The aberrations are introduced in wavefronts when a smooth wavefront is reflected from a non-uniform surface or the propagating wavefront passes through a medium with the non-uniform refractive index [8] (shown in Figure 3.6). The phase of the wavefront is essential while quantifying the aberrations present in the system as the aberrations affect the phase of a wavefront. The phase of the wavefront is defined by

$$\phi = \frac{2\pi}{\lambda} x \tag{3.11}$$

Table 3.1: Representation of different Zernike aberration modes [12].

Index	Noll's Index	Name	Expression	Representation	PSF
$Z_2^{-2}(x,y)$	5	oblique astigmatism	$2xy$		
$Z_2^{2}(x,y)$	6	vertical astigmatism	$x^2 - y^2$		
$Z_3^{-1}(x,y)$	7	vertical coma	$3y^2 + 3x^2y - 2y$		
$Z_3^{1}(x,y)$	8	horizontal coma	$3x^2 + 3y^2x - 2x$		
$Z_3^{-3}(x,y)$	9	vertical trefoil	$3x^2y - y^3$		
$Z_3^{3}(x,y)$	10	horizontal trefoil	$x^3 - 3xy^2$		
$Z_4^{0}(x,y)$	11	primary spherical	$6x^4 + 12y^2x^2 - 6x^2 + 6y^4 - 6y^2 + 1$		

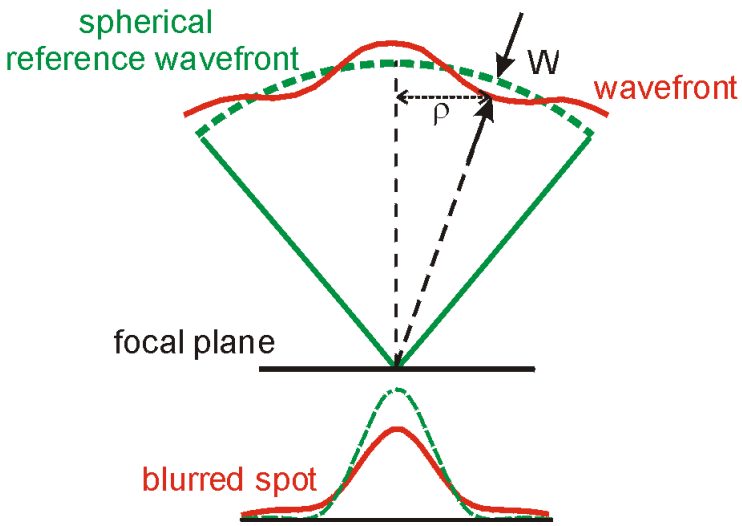


Figure 3.5: Comparison of the effect of undistorted and distorted wavefront on the final image [12].

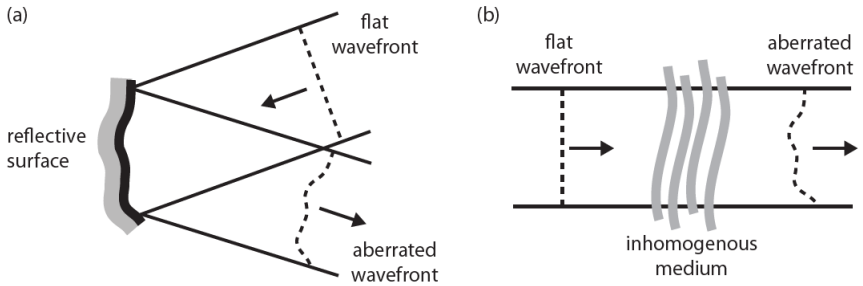


Figure 3.6: (a) Wavefront distortion when light is reflected of a non uniform reflective surface. (b) Wavefront distortion when light passes through a region of non-homogeneous refractive index [41].

where ϕ is the phase of the propagating wave, λ is the wavelength of the wave and x is the optical path length.
Aberrations are incorporated in the definition of the pupil of an optical system. Since most commercial microscopes have a circular pupil, defining the *complex pupil function* on a *unit circle* makes the most sense. The function is defined in the spherical co-ordinate system by

$$P(r,\theta) = \begin{cases} A(r,\theta)e^{\iota\phi(r,\theta)} & 0 \le r \le 1 \\ 0 & r \le 1 \end{cases} \tag{3.12}$$

where $A(r,\theta)$ is the amplitude of the function and $e^{\iota\phi(r,\theta)}$ defines the phase of the *complex pupil function*. The aberrations present in the phase of the function are defined as a set of basis function as shown by the following equation

$$\phi(r,\theta) = \sum_{i}^{N} a_i \psi_i(r,\theta) \tag{3.13}$$

where a_i is the amplitude of a mode and $\psi_i(x,y)$ is the mode. There is an infinite number of choices for basis functions but in microscopy, Zernike polynomials are the used most extensively as they are orthogonal over the interior of a unit circle [12].

Zernike polynomials [75] were introduced by Frits Zernike who invented Phase Contrast Microscopy [11]. The choice of Zernike polynomials has many advantages.The first advantage is that the coefficient of each mode is the RMS (Root Mean Square) wavefront error associated with that mode. Another advantage is that the Zernike coefficient used to describe a wavefront is independent of the number of

polynomials used in the sequence. Zernike modes which have higher magnitude represents a more severe effect on the wavefront and hence degrades the performance of the optical system more. The wavefront can be represented using the Zernike polynomials as

$$W(r,\theta) = \sum_{n,m} C_n^m Z_n^m \tag{3.14}$$

where $W(r,\theta)$ is the wavefront, C_n^m is the RMS magnitude of a particular mode and Z_n^m is the aberration mode. Zernike modes are represented using 2 indices, n representing the radial order and m representing the azimuthal order. The Zernike polynomials can be represented using a complex equation

$$Z_n^m(r,\theta) \pm \iota Z_n^{-m}(r,\theta) = R_n^m(r) e^{\pm \iota m\theta} \tag{3.15}$$

which can be written as

$$\begin{aligned} Z_n^m(r,\theta) &= R_n^m(r) cos(m\theta) \quad m > 0 \\ Z_n^m(r,\theta) &= R_n^m(r) sin(m\theta) \quad m \leq 0 \end{aligned} \tag{3.16}$$

where the radial function $R_n^m(r)$ is defined over a unit circle

$$R_n^m(r) = \sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-l)!}{l![\frac{1}{2}(n+m)-l]![\frac{1}{2}(n-m)-l]!} r^{n-2l} \tag{3.17}$$

A new representation was introduced was introduced by Noll [49] where the Zernike polynomials where defined using a new normalization and with a single index. The table 3.1 shows the representation of some Zernike modes and their blurring effect in the final image. Zernike aberration modes can be categorized into 3 types [66]. The first type of aberration is a mode which does not affect the image quality. Piston (Z_0^0) belongs to the first class of aberrations, which adds a constant phase to the whole wavefront but does not affect the wavefront shape and hence doesn't affect the image quality. Tip (Z_1^{-1}), Tilt (Z_1^1) and Defocus (Z_2^0) belong the second type of aberration in which there is displacement along the x, y and z-axis respectively but the quality of the image is not affected. All the other modes of aberration belong to the third type of aberration which affects the image quality by distorting the shape of the wavefront.

3.3. Fisher Information Matrix and CRLB

The Cramer-Rao lower bound is used as a metric which is used to evaluate the performance of a localization algorithm. The Cramer-Rao lower bound is the lower limit on the precision [56] which can be achieved using an unbiased estimator (an estimator whose estimates results in zero average error). Fundamental properties of an optical system such as photon count, emission wavelength, the numerical aperture of the objective lens, light collection efficiency and image acquisition duration affect the theoretical precision with which a single emitter can be localized. The CRLB depends on the inverse of the square root of the number of photons. Since the acquisition of the data on the detector is a random process as the result of the stochastic nature of the single-molecule emission, the localization problem is statistical. Therefore, the performance of a localization algorithm is characterized as the standard deviation of results obtained from multiple experiments [50][37]. Comparing the CRLB with the 1σ confidence obtained is used to characterize an estimators performance. To compute the CRLB, Fisher Information Matrix is used whose inverse gives the lower bound variance. To calculate the lower bound, the square root of the lower bound variance is computed which is the theoretical limit on the precision of an unbiased estimator. The computation of the lower bound is independent of any estimation method and therefore it serves as a uniform yardstick for comparison. The following equation represents the mathematical formulation of the CRLB :

$$cov(\hat{\theta}) \geq I^{-1}(\theta) \tag{3.18}$$

where $I(\theta)$ is the Fisher Information Matrix, $cov(\hat{\theta})$ is the covariance matrix, $\theta = \{\theta_1, \theta_2, \theta_{\dots}, \theta_n\}$ where θ is the vector of parameters and $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_{\dots}, \hat{\theta}_n\}$ and $\hat{\theta}$ are the estimated parameters. Both the $cov(\hat{\theta})$ and $I(\theta)$ are an $N \times N$ matrix. The inequality does not imply that each element of the $cov(\hat{\theta})$

matrix is bigger than each element of $I(\theta)$ but since the matrix $cov(\hat{\theta}) - I(\theta)$ is positive semi-definite, the main diagonal element of $cov(\hat{\theta})$ is greater or equal to the diagonal element with the same index in the $I(\theta)$ matrix. This relation can be represented mathematically using the following equation:

$$var(\hat{\theta}_i) \geq [I^{-1}(\theta_{ii})] \tag{3.19}$$

where ii is the i^{th} main diagonal element of matrix $I(\theta)$. Any parameter can be estimated from the acquired images such as the signal photon count, background, location of the fluorophores, aberrations present in the optical system as long as the parameters to estimated are appropriately incorporated into the data [13]. The elements of the Fisher information matrix can be defined using the following equation:

$$I(\theta) = E[\frac{\delta ln(L(\theta))}{\delta \theta_i} \frac{\delta ln(L(\theta))}{\delta \theta_j}] \tag{3.20}$$

where $L(\theta)$ is the likelihood function, θ is the set of estimated parameters and E is the expectation. As the emission process is modeled as Poisson process the likelihood function can be written as

$$L(\theta) = \prod_q \frac{\mu_q^{N_q} e^{-\mu_q}}{N_q!} \tag{3.21}$$

where N is observed data, μ is the expectation model and q represents the pixel index. Using this definition of the likelihood function the Fisher Information Matrix can be rewritten as

$$I(\theta) = \sum_q \frac{1}{\mu_q} \frac{\delta \mu_q}{\delta \theta_i} \frac{\delta \mu_q}{\delta \theta_j} \tag{3.22}$$

4

smNet Architecture and Workflow

In this section, the working and the architecture of the deep learning method smNet which forms the base of this research are discussed. The deep neural network, smNet proposed by Zhang *et al* [76] can be used to perform 3D localization, dipole orientation estimation and wave-front aberration.

4.1. smNet Architecture

Zhang et al present a deep learning architecture for 3-D localization, orientation estimation and wave-front aberration where each of the problems is modelled as a supervised regression problem. Multiplexing of the training process for each task reduces the complexity of the whole problem. Using a single network to perform 3D localization, orientation estimation and wavefront aberration would make the final complexity multiplicative. This would result in huge parameter space and the training process would require an incredible amount of training data. By multiplexing the tasks, the complexity of the problem is made additive, therefore, resulting in the reduction of the parameter space. The reduction of the parameter space also means that the amount of training data required is also reduced significantly. The input to the smNet (see figure 4.1 and tables 4.1 and 4.2) is a 3D image with the dimensions $C \times N \times N$ pixels (image size N = 16 or 32) where C stands for the number of channels in the input image (C = 1 for single plane PSF and C = 2 for biplane PSF, for example). The network of the pipeline which performs 3D localization and orientation estimation is 28 layers deep and contains 5 convolution layers followed by batch normalization layers after each convolution layer, 7 residual layer (each residual layer contains 3 convolution layers) and 2 fully connected layers. The network for aberration detection pipeline which estimates up to 12 Zernike coefficients is similar to the 3D localization network and the only difference in network architecture is that aberration detection network has only 1 fully connected layer instead of 2 layers. The network used for aberration detection with more than 12 Zernike modes up to 21 Zernike modes has 2 convolution layers, 11 residual layers and batch normalization layers after each convolution layers. For the task of xy localization, z localization and angle estimation, the multiplexing is done while calculating the error by making selection of the appropriate cost function used to train the network. While performing aberration estimation, multiplexing is done at an architectural level apart as well as while calculating the error. The figure represents the architecture of the network which is used to perform 3D localization and aberration estimation. The grey rectangles represent the convolution layers, the coloured rectangles represent the residual layer and the last 2 layers represent the fully connected layers. Table 4.1 describes the detailed architectural information of each layer of the network pipeline which performs aberration estimation. Table 4.2 describes the detailed architectural information of the network which is used to estimate the up to 12 modes of wavefront aberration.

Convolution Layer

Convolution layers [33] are extremely efficient in extracting spatial information and are therefore extensively used in image-based applications such as object detection, image classification. A convolution layer takes an image as input of a size C×N×N where C is the number of the channel of the image and N is the number of pixels. It generates feature maps by convolving kernels whose weights are

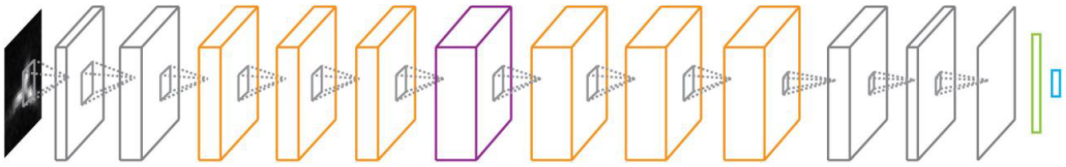


Figure 4.1: Visual representation of the smNet architecture.

Table 4.1: Details of smNet architecture used to perform 3D localization.

Building Blocks	Kernel Size	Stride	Output Size
conv(1)	C×7×7	1	64×N×N
conv(2)	64×5×5	1	128×N×N
res(1-3)	$\begin{bmatrix} 128\times3\times3 \\ 32\times3\times3 \\ 64\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 32\times N\times N \\ 64\times N\times N \\ 128\times N\times N \end{bmatrix}$
res(4)	$\begin{bmatrix} 128\times3\times3 \\ 64\times3\times3 \\ 128\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 64\times N\times N \\ 128\times N\times N \\ 256\times N\times N \end{bmatrix}$
res(5-7)	$\begin{bmatrix} 256\times3\times3 \\ 64\times3\times3 \\ 128\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 64\times N\times N \\ 128\times N\times N \\ 256\times N\times N \end{bmatrix}$
conv(3)	256×1×1	1	128×N×N
conv(4)	128×1×1	1	64×N×N
conv(5)	64×5×5	1	1×N×N
FC(1)	—	—	10
FC(2)	—	—	1 or 2

initialized randomly. Each feature map contains some information about the input images. The higher the number of feature maps generated the more spatial information about the image is collected. So, smNet takes a 1 ×N ×N image and generates 64×N ×N output with a kernel size of 7 in the first layer. It means in the first layer of smNet 64 kernels of size 7×7 are generated with random weights. These randomly initiated kernels are used to perform convolutions and generate 64 feature map of the size N ×N. The initial feature maps are used to extract low-level information such as vertical edges and horizontal edges. The feature maps in the subsequent layers extract high-level information from the feature maps generated in the previous layers. While training, the weights of the randomly initialized kernels are adjusted so that the error in prediction coming out of a network and ground truth is minimized. Strides in the convolution layers are used to control the output size of the feature maps. If the stride is 2, then the feature map generated will be sub-sampled by a factor of 2. The figure 4.2 shows a visual representation of the process of generation of feature maps in a convolution layer.

Residual Layer

A common problem while training very deep neural networks is the problem of exploding or vanishing gradient. To tackle this, the concept of residual blocks [21] or skip connection was introduced. The figure 4.3 shows how a skip connection can be implemented. It was observed that without skip connections the training error first decreases and after a certain number of layers the training error starts increasing. When a skip connection is implemented the training error keeps decreasing with the increase in the number of layers. Each unit with a skip connection as shown in the figure 4.3 is called a residual block. Such residual blocks are stacked one after the other to train deep neural networks.

The following equations represents the mathematics of the residual blocks

$$\begin{aligned} X^{r+1} &= \sigma(R(w^r, X^r) + f(w_f^r, X^r)) \\ R(w^r, X^r) &= BN(F_{ca}(w^{r_{l3}}, \sigma(BN(F_{ca}(w^{r_{l2}}, \sigma(BN(F_{ca}(w^{r_{l1}}, X^r)))))))) \end{aligned} \tag{4.1}$$

Table 4.2: Details of smNet architecture used to perform aberration estimation.

Building Blocks	Kernel Size	Stride	Output Size
conv(1)	C×7×7	1	64×N×N
conv(2)	64×5×5	1	128×N×N
res(1-3)	$\begin{bmatrix} 128\times3\times3 \\ 32\times3\times3 \\ 64\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 32\times N\times N \\ 64\times N\times N \\ 128\times N\times N \end{bmatrix}$
res(4)	$\begin{bmatrix} 128\times3\times3 \\ 64\times3\times3 \\ 128\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 64\times N\times N \\ 128\times N/4\times N/4 \\ 256\times N/4\times N/4 \end{bmatrix}$
res(5-7)	$\begin{bmatrix} 256\times3\times3 \\ 64\times3\times3 \\ 128\times3\times3 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 64\times N/4\times N/4 \\ 128\times N/4\times N/4 \\ 256\times N/4\times N/4 \end{bmatrix}$
conv(3)	256×1×1	1	128×N/4×N/4
conv(4)	128×1×1	1	64×N/4×N/4
conv(5)	64×5×5	1	1×N/4×N/4
FC(1)	—	—	12

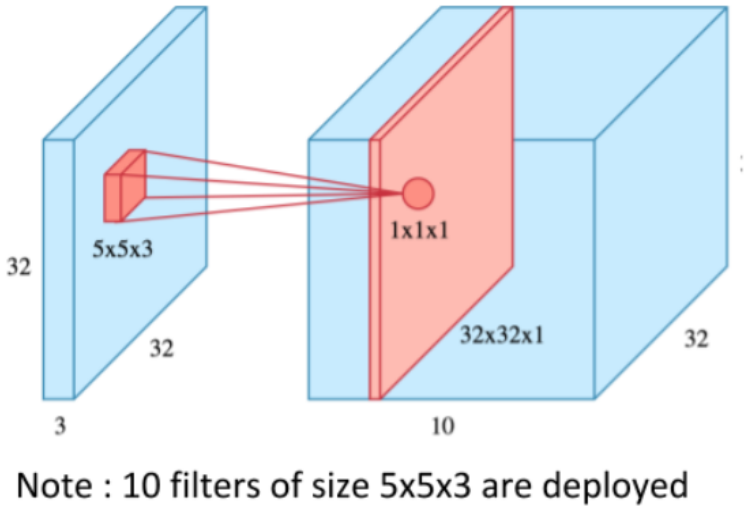


Figure 4.2: Generation of a feature map in CNN

$$f(w_f^r, X^r) = \begin{cases} X^r & p = q \\ BN(F_{c1}(w_f^r, X^r) & p \neq q \end{cases} \tag{4.2}$$

where σ represents the PReLU activation function, BN represents Batch Normalization layer, w^r and w_f^r are parameters of the residual block and identity function f. X^r and X^{r+1} are the input and output of a residual block respectively. F_{ca} is a 3×3 convolution and $w^{r_{l1}}$, $w^{r_{l2}}$, $w^{r_{l3}}$ are parameters to the convolution layers r_{l1}, r_{l2} and r_{l3}. F_{c1} is a 1×1 convolution and p and q are the number of feature maps in the input and output layers of a residual blocks, respectively, The identity function in the residual block enables the back-propagation signal to reach the input layer from the output layer in a deep neural network.

Fully Connected Layer

Fully connected layers are powerful feature extractors. Fully connected layers work on the universal approximation theorem [7] which states that a neural network with a finite number of neurons and non-linear activation is capable of predicting any real-valued function. Still, fully connected layers are not

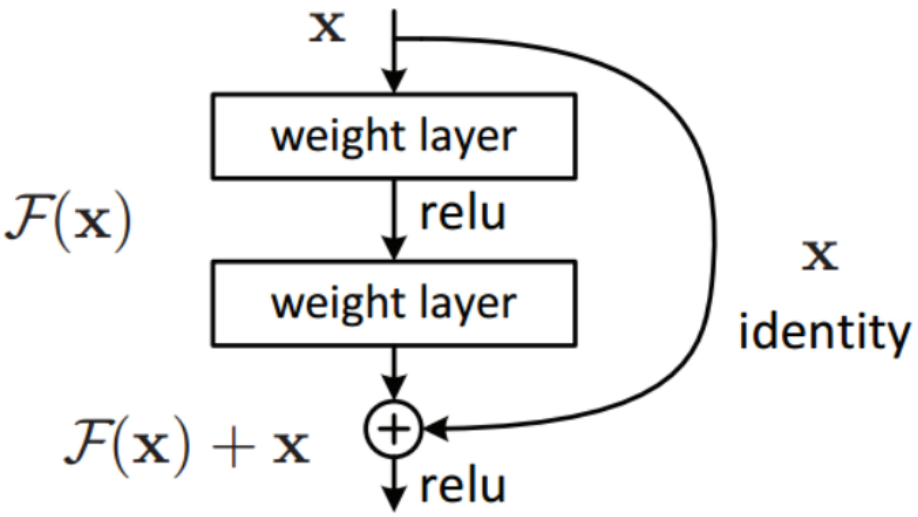


Figure 4.3: Implementation of skip connections in residual blocks.

used to create deep networks because even a few fully connected layers significantly scale up the number of trainable parameters and make the model large and computationally expensive. It is the limitation on the available computational power which limits the usage of fully connected layers in deep learning. Generally, it is used in the final layer of a deep neural network where a real-valued outcome is expected.

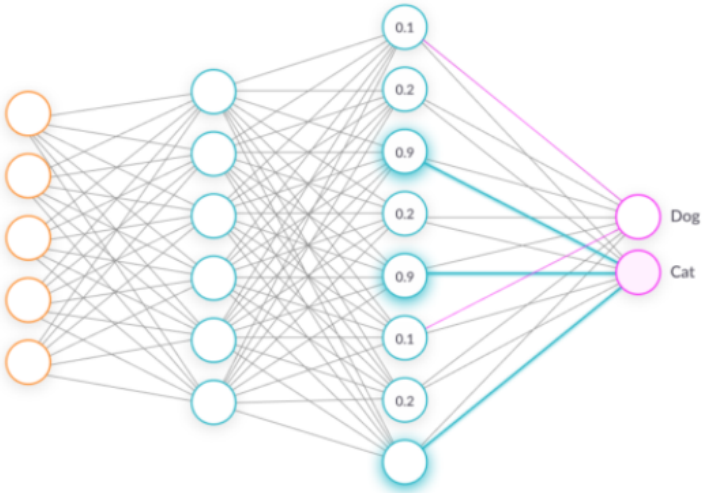


Figure 4.4: Representation of fully connected layers in a neural network.

Batch Normalization Layer

For training the neural networks, if the weights are updated after passing a single dataset through the network the training process would become noisy. Instead, training is done in batches where the average error of a single batch is used to update the weights. Batch normalization is used to between layers as a measure to counteract internal covariate shift [27] by normalizing the gradients coming out of a layer and prevents the problem of exploding gradients. It is also seen that adding batch normalization layer makes the training process significantly faster. The training process becomes faster as all the features after batch normalization have a mean of zero and a standard deviation of one and it makes

the parameter space smaller. In some cases, it is also seen that adding batch normalization layers increases the accuracy of the model. Since the normalization step takes place when the error for a batch is computed the layer is called batch normalization layer. In smNet, the batch normalization layer is present after each of the convolution layers and it prevents the problem of exploding gradients.

Activation Functions

Activation functions are used to introduce non-linearity in the neural networks. With the non-linearity in the system, the neural networks would essentially perform linear regressions. In smNet, parameterized ReLu (PReLu) or leaky ReLu have been used after each CNN and residual layers as it does not saturates. The figure shows the PReLu function which is defined by the following equations :

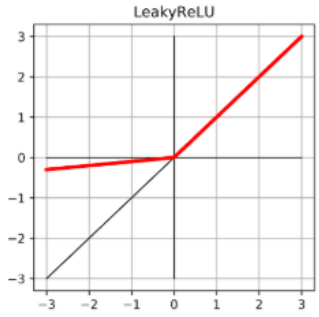


Figure 4.5: This figure shows the leaky ReLu or Paramterized Relu (PReLu) function.

$$f(x) = \begin{cases} ax & x < 0 \\ x & x \geq 0 \end{cases} \tag{4.3}$$

where a is a tunable hyperparameter. In smNet, the default initialization provided by PyTorch is used. The final fully connected layer of the model which is used for 3D localization and orientation estimation uses HardTanh activation. The HardTanh activation (shown in the figure) is represented by the following equation.

Figure 4.6: This figure shows the HardTanh function.

4.2. smNet training algorithm

The figure 4.7 shows the steps involved in training smNet. The dataset is normalized, divided into training and validation data and training data is passed through the network to generate output. Using the output the error is computed and the weights are adjusted to reduce the error. Once after every pass, the validation data is passed through the network to evaluate the validation error which gives insight about the networks ability to generalize.

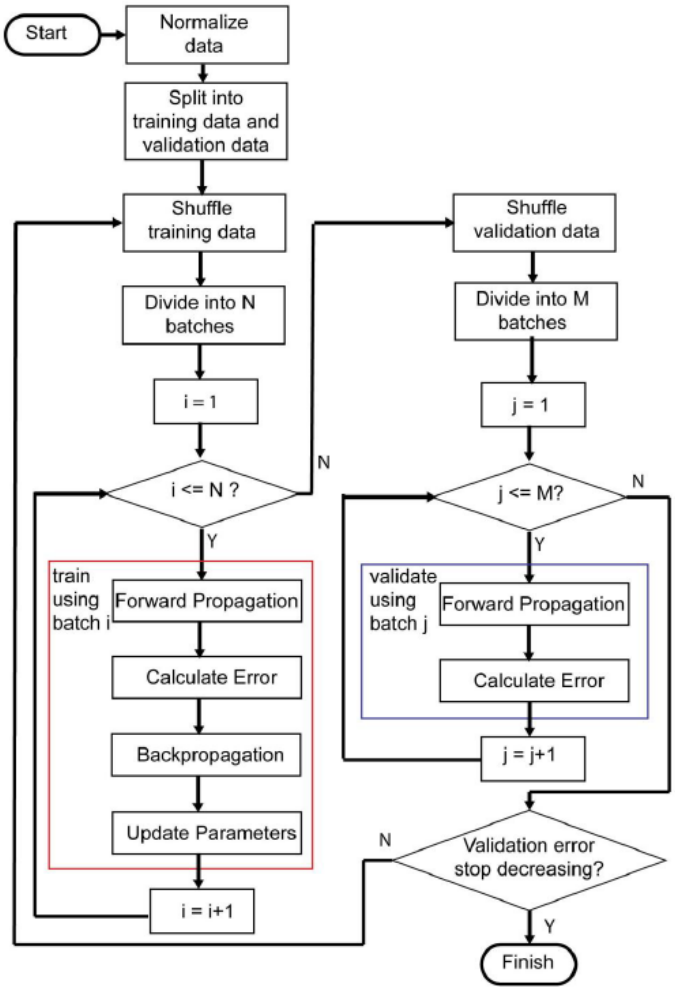


Figure 4.7: Flowchart of the smNet training process

Training and Validation Split

smNet is a supervised learning algorithm where the images are fed to the neural network with the ground truth data. While training smNet to perform 3D localization, the ground truth data or the labels are the spatial locations of the single-molecule emitters. While the network is trained to perform angle orientation, the labels are the azimuthal and polar orientation of the single-molecule emitter. To perform wavefront aberration estimation, while training the network the labels are the Zernike coefficient of each mode of aberration present in the optical system. The training data and the validation data is split in a 75 % to 25 % ratio and all the inputs are normalized and fed to the neural network for training.

Weight and Bias Initialization

The use of batch normalization has many benefits. It enables using of a higher learning rate and reduces the dependence of the training process on the weight initialization [27]. So the weights and bias of each of the filter kernels in the convolution layers and the connections in the fully connected layer was generated from a uniform distribution with a range of $[-stdv,-stdv]$ where $stdv$ is defined as

$$stdv = \begin{cases} \frac{1}{\sqrt{kW \times kH \times N_{inputplanes}}} & convolution \\ \frac{1}{InputSize} & fullconnected \end{cases} \tag{4.4}$$

where kW and kH are the height and width of the filter kernel and $N_{InputPlane}$ is the information about the number of input channels.

Forward Propagation

Forward propagation is used to generate the output of a neural network depending upon the existing weights and biases present in the network. The final output is an estimate of either the 3D positions, orientation angles or the wavefront aberrations. For generating the output of each pass, the training data is divided into batches of 128 images and an estimate of each batch is generated. When all the batches produce an output it is called an epoch. Forward propagation of smNet can be mathematically represented by :

$$A_i^{l+1} = \begin{cases} \sigma(BN(F^l(w^l, A_i^l))) & convolutional \\ \sigma(R(w^l, A_i^l) + f(w_f^l, A_i^l)) & residual \\ \sigma((F^l(w^l, A_i^l))) & fully-connected \\ HT((F^l(w^l, A_i^l))) & fully-connected(HardTanh) \end{cases} \tag{4.5}$$

where A_i^{l+1} represents the output of a layer for for the mini batch i for the layer l. BN represents the Batch Normalization operation, σ represents the PReLu non-linear activation function and HT represents the HardTanh non-linear activation function. F^l represents the linear transformation applied on the input of a previous layer and f is the identity function of a residual block. The outputs generated by the various pipelines of smNet after a forward pass are :

$$A_i^l = \begin{cases} (\hat{x}, \hat{y}) \\ \hat{z} \\ \hat{\alpha} \\ \hat{\beta} \\ (\hat{Zer}_1, \hat{Zer}_2, ...) \end{cases} \tag{4.6}$$

Error Calculation

Once the network generates an estimate after the forward pass, the difference between the ground truth and the network estimate is calculated which is known as the error. The Error is computed in the form of the cost function and the neural network tries to reduce the cost function by adjusting the weights and biases to make the predictions as close as possible to the ground truth. The cost function used in smNet for calculating the error is a CRLB-weighted minimum square loss function. The CRLB weighting makes the algorithm try to improve the precision compared to the theoretical limit. The error is computed for each mini-batch and the average error of all the mini-batches is used to adjust the weights and biases. The cost function used to calculate the error in smNet is described using the equation 4.7. The training error continuously decreases with each passing epoch and the training error is stopped when the validation error converges and doesn't change for a few epoch. It is defined by :

$$E_{\hat{\theta}} = \frac{1}{NT} \sum_{n=1}^{N} \sum_{t=1}^{T} \frac{(\hat{\theta}_{tn} - \theta_{tn})^2}{CRLB_{\theta_{tn}}} \tag{4.7}$$

where N is the number of images in each batch, T is the size of the output, $\hat{\theta}$ being the networks estimate and θ is the ground truth.

Weight Updation

Once the average error is computed in each epoch the weights and biases of the networks are updated using the Adam optimizer. The Adam optimizer is based on a modified implementation of the stochastic gradient descent [28]. In stochastic gradient descent, the learning rate is constant and this might lead to the error not converging. The Adam optimizer uses an adaptive learning rate which ensures convergence. Furthermore, it uses an exponential moving average of the gradient (if the gradient descent is moving in the correct direction towards the minima the learning rate is increased by taking

a moving average of the gradients) and the squared gradient. The gradient-based weight updation process is defined by :

$$w_{k+1}^{l} = w_{k}^{l} - \frac{\eta}{M} \sum_{n=1}^{M} \frac{\partial E_{\hat{\theta_{n,k}}}}{\partial w_{n,k}^{l}} \tag{4.8}$$

where w_k are the existing weights and w_{k+1} are the updated weights, η is the initial learning rate and k is the iteration number and M is the number of images in a mini batch.

Stopping Point

During the training process, the neural network can learn noise and useless features. This can cause the performance of the neural network to degrade when it is used to make predictions on the test dataset. This degradation in the network's performance is called overfitting and the validation dataset is used to measure the network's generalization capability. The training error continuously decreases while the training error converges at a certain point. The training process is stopped empirically when the validation error converges. For smNet when the validation error is equal to one it means that smNet had learnt all possible information present in the image and the performance of the smNet is equal to the theoretical limit.

Regularization: Batch Normalization vs Dropout

Dropout is a popular method which is used to reduce overfitting in neural networks. This is done by randomly switching off a few connections while training. This leads to an ensemble type of learning where each of the updates is done using a different configuration. In smNet, it was found that introducing dropout does not reduce overfitting. The possible reason was that smNet uses very small images for training the network and randomly dropping connections may reduce the spatial resolution and make the smNet's performance unstable [26]. Batch Normalization was used as a method to reduce overfitting and it was found that adding a Batch Norm layer after the convolution layer improves the performance of smNet.

4.3. smNet workflow

When the smNet models are trained they can be used to make predictions on experimental data. The steps involved in making a prediction using an smNet network on an image generated by a fluorescence microscope are :-

- Channel Registration (For multi plane PSFs)
- Segmentation
- smNet estimation
- Rejection
- Averaging

Channel registration is an essential step to use smNet with microscopes acquiring bi-plane PSFs. The rigid registration step aligns the images obtained in the 2 planes. The affine transformation is computed on a couple of beads and the transform is used to register the images. After the registration is done, candidate emitters are selected from the whole field of view. The selection of candidate emitters is done by identifying the local maxima over a user-defined threshold present in the image. A 16×16 region or 32×32 region is segmented from the image keeping the local maxima as the centre. The size of the segmented image depends on the size of the training data used to train smNet. The trained smNet model is used to predict the axial and lateral position, dipole angle or the wavefront aberration which is present in the optical system. Since this localization information is used by reconstruction algorithms to generated super-resolved images, only localizations which can be reliable are retained. The rejection of unreliable localization is done by estimating the photon count and background present in a segmented

image and computing the SNR. If the SNR is below a user-defined threshold, the predictions of smNet are rejected. To ensure the generation of accurate images of the biological structures, the localization needs to be robust. This is ensured by averaging the estimates from multiple images obtained from the microscope from the same field of view.

5

Methods

In this section, the design of all the simulation experiments which were carried out are discussed. In addition to the simulation experiments, experiments are performed on experimentally collected data used by Thorsen *et al* in their work 'Impact of optical aberrations on axial position determination by photometry [67].' All the simulation experiments comprised of three parts. The first part of the simulation experiments was the generation of training data. To generate the training data, the Vector 3D PSF model [59][63] was used. The training data was generated on the local computer running MATLAB R2019a. The local computer was running on Intel i7-8750H processor (12 cores) with a clock speed of 2.2 GHz. It had 16 GB of RAM and 4 GB Nvidia Quadro P1000 graphics card. The second part of the simulation experiments was to train the neural network. For training the neural network, a high-performance cluster was used with 32 CPUs, 192 GB memory and 16 GB GPU. For training the neural network, Pytorch 0.4.0 and CUDA 10.1 libraries were required. The last part of the simulation experiments was to generate test data and test the performance of the neural network which was done on the local machine. The various simulation experiments are designed to characterize the performance of smNet trained on simulated images. The performance of smNet is characterized over a large range of varying physical conditions such as signal photon count, background count, aberration intensity and aberration modes. Once the characterization of smNet's performance is done, a pipeline is developed which can be used to efficiently train the smNet to deliver robust performance irrespective of physical conditions. The concept of simulator learning is also tested using smNet where a neural network is trained with purely simulated data and it is used on experimental data without any retraining. The performance of smNet trained on simulated data is compared to the performance of a state of the art fitting based algorithm [67] on experimentally obtained data.

5.1. Characterization of the performance of smNet with an accurate PSF model

Lateral localization of single emitters simulated at the focus - constant photon and background count

The first simulation experiment was done to characterize the lateral localization performance of smNet when the signal photon count and the background count were constant and equal in both the training and the test data. In this experiment, an smNet model was trained to perform lateral localization of a single emitter which was at the focus. Figure 8.1 (Appendix A) is a representation of the training data which was used to train the neural network and figure 8.5 (Appendix A) represents the training curve which was obtained while training the neural network. The parameters which were used to train the neural network are listed in the table 5.1.

Table 5.1: Experiment 1 - Training smNet to perform lateral localization at the focus with constant signal photon and background

Localization Mode	xy localization
Training Data Size	16x16x1x100000
Test Data Size	16x16x1x1000
x and y range	[-5 pixel 5 pixel]
z range	0
Signal Count	2000 photons
Background Count	20
Aberration Mode	None
Tanh Limit	6
Batchsize	128

Axial and lateral localization of single emitters with defocus - constant signal photon and background count

In the second simulation experiment, the motivation was to characterize the performance of smNet in performing 3D localizations where the signal photon count, background count and vertical astigmatism were constant and equal in both training and test data. In this simulation experiment, smNet was trained to perform both axial and lateral localization of a single emitter where the single emitter could be randomly present in the axial plane between -500 nm to 500 nm. In this simulation, vertical astigmatism was introduced to break the symmetry of the PSF shape in the axial plane above and below the focus. Figure 8.2 (Appendix A) shows a representation of the training and the test data used. The neural network is trained separately for xy and z localization and both the training curves are presented in the figure 8.6 (Appendix A). The table 5.1 presents information about the parameters which were used to train the neural network.

Table 5.2: Experiment 2 - Training smNet to perform 3D localization with constant signal photon, background and aberration.

Localization Mode	xyz localization
Training Data Size	16x16x1x120000
Test Data Size	16x16x1x1000
x and y range	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]
Signal Count	2000 photons
Background Count	20
Aberration Mode	[2,2]
Aberration Level	72 mλ
Tanh Limit - xy	4
Tanh Limit - z	2
Batchsize	128

Axial and lateral localization of single emitters with defocus - signal photon count and background dependence

The next simulation experiment was done to characterize the ability of smNet in performing 3D localization when the test images had varying level of signal photon count and background count. The level of astigmatism used to break the symmetry of PSFs above and below the focus was kept constant in both the training and the test images. In this simulation experiment, the smNet model was trained with images with the signal photon and background count varying as Gaussian distributions (details in the table 5.3). To characterize the performance of the smNet as a function of the signal photon, multiple test data set were generated. Each test data set had 1000 images where the background was constant (20) and the signal photon count was of a certain value. The performance of the smNet was characterized over a range a large range of photon count which varied from 1000 photons to 8000 photons. A similar design was chosen to characterize the performance of smNet as a function of background. Multiple

test data set was generated where all the images had a signal photon count of 4000. Each set of 1000 images had a certain level of background. The background was varied from 10 to 200. The figure 8.7 (Appendix A) shows the training curve which was obtained while training the network where the signal photon count and the background count was sampled randomly from a Gaussian distribution.

Table 5.3: Experiment 3 - Characterizing the 3D localization performance of smNet as a function of signal photon and background

Localization Mode	xyz localization
Training Data Size	16x16x1x120000
Test Data Size	16x16x1x1000
x and y range	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]
Signal Count	Gaussian Distribution, μ = 4000, σ = 1200
Background Count	Gaussian Distribution, μ = 25, σ = 15
Aberration Mode	[2,2]
Aberration Level	72 mλ
Tanh Limit - xy	4
Tanh Limit - z	2
Batchsize	128

Axial and lateral localization of single emitters with defocus - dependence on aberration intensity

This simulation experiment was done to characterize smNet's performance in performing 3D localization in images which had a different aberration intensity compared to the simulated images which were used to train the network. This experiment was done to find out how capable smNet was in dealing with aberration intensity levels which it hasn't encountered in the training process. For this experiment, 2 models were trained where each model was trained with a certain level of fixed vertical astigmatic aberration (see table 5.4 for more details). Test data sets were generated similarly and 1000 images were generated for each signal photon level which was varied from 1000 photons to 8000 photons. Generating test data set with varying level of photon count was used to characterize the performance of the networks over a range of signal photons. Figure 8.3a and 8.3b (Appendix A) show the representation of the PSFs used to train the 2 different model with simulated data having vertical astigmatism of 36 $m\lambda$ and 72 $m\lambda$ respectively.

Table 5.4: Experiment 4 - Characterizing the performance of smNet on test data having different aberration intensity

Data	36 mλ	**Data**	72 mλ
Localization Mode	xyz localization	**Localization Mode**	xyz localization
Training Data Size	16x16x1x120000	**Training Data Size**	16x16x1x120000
Test Data Size	16x16x1x1000	**Test Data Size**	16x16x1x1000
x and y range	[-3 pixel 3 pixel]	**x and y range**	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]	**z range**	[-500 nm 500 nm]
Signal Count	μ = 4000, σ = 1600	**Signal Count**	μ = 4000, σ = 1600
Background Count	μ = 25, σ = 5	**Background Count**	μ = 25, σ = 5
Aberration Mode	[2,2]	**Aberration Mode**	[2,2]
Aberration Level	36 mλ	**Aberration Level**	72 mλ
Tanh Limit - xy	4	**Tanh Limit - xy**	4
Tanh Limit - z	2	**Tanh Limit - z**	2
Batchsize	128	**Batchsize**	128

Axial and lateral localization of single emitters with defocus - 1 aberration mode vs 5 aberration mode

To characterize the performance of smNet in localizing single emitters in the presence of different modes of wavefront aberrations, this simulation experiment was carried out. This was done by training 2 smNet models. The first model was trained with images having 1 mode of wavefront aberration (see to the table 5.5). The second model was trained with images having 5 modes of wavefront aberration present in the simulated optical setup. To test the performance of these networks, multiple test sets were generated. A test data set was generated with only one mode of aberration and the test data set had images with varying level of signal photon count to test the performance of both the trained network as a function of signal photon count. Similarly, a test data set was generated in which 5 modes of wavefront aberration were present. The performance of both the model was evaluated by running both the models on the two test data set. The performance of both the models was tested by varying the signal photon count from 1000 photons to 8000 photons.

Table 5.5: Experiment 5 - Characterizing the performance of smNet on test data having different aberration modes

Data	1 Aberration Mode	**Data**	5 Aberration Mode
Localization Mode	xyz localization	**Localization Mode**	xyz localization
Training Data Size	16x16x1x120000	**Training Data Size**	16x16x1x120000
Test Data Size	16x16x1x1000	**Test Data Size**	16x16x1x1000
x and y range	[-3 pixel 3 pixel]	**x and y range**	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]	**z range**	[-500 nm 500 nm]
Signal Count	μ = 4000, σ = 1600	**Signal Count**	μ = 4000, σ = 1600
Background Count	μ = 25, σ = 5	**Background Count**	μ = 25, σ = 5
Aberration Mode	[2,2]	**Aberration Mode**	[2,-2],[2,2],[3,1],[3,-1],[4,0]
Aberration Level	72 mλ	**Aberration Level**	72 mλ
Tanh Limit - xy	4	**Tanh Limit - xy**	4
Tanh Limit - z	2	**Tanh Limit - z**	2
Batchsize	128	**Batchsize**	128

Axial and lateral localization of single emitters with defocus - splitting of the parameter space

In a step towards simulator learning, an smNet model should perform robust 3D localizations irrespective of the level of aberration present in the optical system. To characterize the performance of smNet in dealing with random signal count, background, aberration intensity, aberration modes, 4 smNet models are generated where each model is trained with a certain root mean square (RMS) level of wavefront aberration (refer table 5.6 for more information) comprising of oblique astigmatism, vertical astigmatism, vertical coma, horizontal coma and primary spherical aberration. The random aberrations of a certain RMS value which are generated are added to a constant level of vertical astigmatism which is used to break the symmetry in PSF shapes in the axial plane above and below the focus. Four sets of test data are generated similarly and the performance of all the models are tested on all the four test data sets.

Aberration Estimation on Simulated Data

To characterize the performance of smNet in estimating wavefront aberration, a model is trained with simulation data containing various levels of wavefront aberration. The network is trained to estimate the first 5 aberration modes: oblique astigmatism, vertical astigmatism, vertical coma, horizontal coma and primary spherical aberration. The aberration coefficient for each of the Zernike mode is generated randomly from the volume of a 5-dimensional hypersphere where the surface of the hypersphere represents the RMS value of 150 mλ. Randomly a W_{RMS} is chosen and then random coefficients are chosen whose RMS value is equal to W_{RMS} and that wavefront aberration is added to an image. This process is done repeatedly to generate the training and test data. The parameters used to generate the training and test data is described in the table 5.7.

Table 5.6: Experiment 6 - Splitting of the parameter space

Data	RMS level 1	**Data**	RMS level 2
Localization Mode	xyz localization	**Localization Mode**	xyz localization
Training Data Size	16x16x1x160000	**Training Data Size**	16x16x1x160000
Test Data Size	16x16x1x1000	**Test Data Size**	16x16x1x1000
x and y range	[-3 pixel 3 pixel]	**x and y range**	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]	**z range**	[-500 nm 500 nm]
Signal Count	μ = 4000, σ = 1600	**Signal Count**	μ = 4000, σ = 1600
Background Count	μ = 25, σ = 5	**Background Count**	μ = 25, σ = 5
Aberration Mode	[2,-2],[2,2],[3,1],[3,-1],[4,0]	**Aberration Mode**	[2,-2],[2,2],[3,1],[3,-1],[4,0]
Aberration Level	0 mλ	**Aberration Level**	36 mλ
Constant [2,2]	54 mλ	**Constant [2,2]**	54 mλ
Tanh Limit - xy	4	**Tanh Limit - xy**	4
Tanh Limit - z	2	**Tanh Limit - z**	2
Batchsize	128	**Batchsize**	128

Data	RMS level 3	**Data**	RMS level 4
Localization Mode	xyz localization	**Localization Mode**	xyz localization
Training Data Size	16x16x1x160000	**Training Data Size**	16x16x1x160000
Test Data Size	16x16x1x1000	**Test Data Size**	16x16x1x1000
x and y range	[-3 pixel 3 pixel]	**x and y range**	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]	**z range**	[-500 nm 500 nm]
Signal Count	μ = 4000, σ = 1600	**Signal Count**	μ = 4000, σ = 1600
Background Count	μ = 25, σ = 5	**Background Count**	μ = 25, σ = 5
Aberration Mode	[2,-2],[2,2],[3,1],[3,-1],[4,0]	**Aberration Mode**	[2,-2],[2,2],[3,1],[3,-1],[4,0]
Aberration Level	72 mλ	**Aberration Level**	104 mλ
Constant [2,2]	54 mλ	**Constant [2,2]**	54 mλ
Tanh Limit - xy	4	**Tanh Limit - xy**	4
Tanh Limit - z	2	**Tanh Limit - z**	2
Batchsize	128	**Batchsize**	128

5.2. Design of Pipeline for Simulator Learning

After characterizing the performance of smNet over a broad range of physical conditions, three hypotheses were developed. The first hypothesis was “Deploying multiple smNet models each designed to tackle a small parameter space, precise and accurate 3D localizations could be performed over a large parameter space”. The second hypothesis which was developed was “Splitting the parameter space, the training process of smNet could be optimized. The required training data and computational load could be reduced without affecting the performance of smNet”. To test these hypotheses, a pipeline was designed which is shown in figure 5.1. The proposed pipeline takes a region of interest (ROI) containing a single emitter as an input. First, the aberration level is estimated using a smNet model trained to estimate wavefront aberrations present in an image. Once an estimate about the W_{rms} (RMS wavefront intensity) is made, model selection is performed. In model selection, out of the many smNet models trained to perform 3D localization, the model which is trained with simulated images having aberration intensity closest to the estimated aberration intensity is chosen to perform 3D localization of the single emitter. In this way, the problem is made additive instead of multiplicative in covering the whole parameter space and would, therefore, reduce the amount of training data which is required to train the localization model. To test this hypothesis, an experiment is designed where a test dataset having 1000 images with a certain level of aberration present is fed to the aberration estimation model. The estimates from the model are used to select the appropriate localization model. Then, localization is performed with the selected model and with a model trained with wrong aberration level which acts as a control. The null hypothesis would be “There is no significant difference in the localization accuracy and precision when localization is performed with the selected model and control model.”

Table 5.7: Experiment 7 - Training smNet to perform aberration estimation

Localization Mode	Aberration Estimation
Training Data Size	16x16x1x1600000
Test Data Size	16x16x1x1000
x and y range	[-3 pixel 3 pixel]
z range	[-500 nm 500 nm]
Signal Count	Gaussian Distribution, μ = 4000, σ = 1600
Background Count	Gaussian Distribution, μ = 25, σ = 5
Aberration Mode	[2,2],[2,-2],[3,-1],[3,1],[2,-2],[4 0]
Aberration Level	[0 mλ 150 mλ]
Batchsize	128

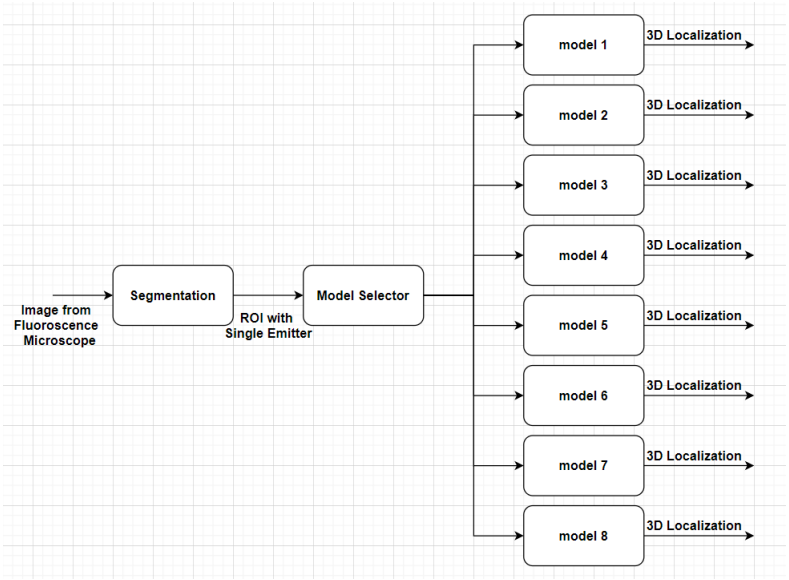


Figure 5.1: Proposed workflow for smNet to perform 3D localization using the concept of Simulator Learning

5.3. Simulator Learning

The third hypothesis which was developed was “Since the accurate PSF model replicates the reality very closely, smNet trained on simulated images can be used directly to make estimations on experimentally obtained images without any retraining”. To test this hypothesis, a smNet model which was trained to perform aberration estimation on simulated images was deployed to estimate the wavefront aberration present in experimentally obtained images. Wavefront aberration estimation was also performed using a very accurate vector fitting based algorithm [67] which was used to verify the level of aberration present in the experimental data which is generated by PSF engineering using an SLM in the optical path [67]. The null hypothesis was “The smNet model trained on simulated data cannot be used directly on experimental data without retraining the model.”

6

Results and Discussion

In this section, the results of the various simulation experiments which were performed are discussed. In the first section, the performance of the smNet in doing 3D localization and aberration estimation is characterized. In the second section, the performance of the proposed pipeline is evaluated. Lastly in the third section, the concept of simulator learning is tested using smNet and its performance is compared with the vector fitter algorithm [67].

6.1. Performance characterization of smNet

Characterization of localization performance

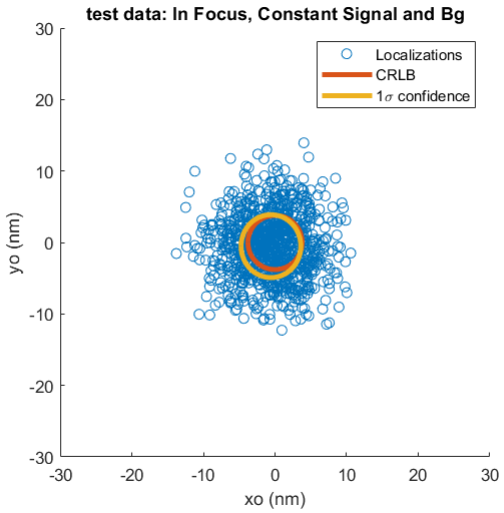


Figure 6.1: Scatter plot of localizations for in-focus single emitter molecules when the signal photon count and background count is constant.

The first experiment was done to characterize the performance of smNet in localizing single-molecule emitters at the focus along the x-y axis when the signal photon count and the background were constant (details in Chapter 5). This was done to find out how accurate and precise smNet was in performing the simplest localization task when it was trained with the accurate vector model. The images were simulated using a wavelength of 690 nm, the NA of the objective lens was 1.4 and the effective pixel size was 100 nm (slight oversampling as the Nyquist rate was 123 nm). The imaging conditions were kept the same in all the experiments. To characterize the performance of smNet in doing lateral localization of single-molecule emitters at the focus, a smNet model was trained with simulated images where the signal photon count and background were kept constant. Figure 6.1 shows the performance of smNet

in localizing 1000 single emitters in the x-y plane. The computed Cramer-Rao lower bound for both x and y-axis were 3.78 nm. The experiment was repeated 10 times with each experiment having 1000 localizations. The performance of smNet trained with the vector model was similar to its counterpart trained with the diffraction model. Similar to the diffraction model, negligible bias was observed. Along the x-axis, the bias was -0.26 nm $\pm$ 0.15 nm (mean bias $\pm$ std) and along the y-axis, the bias was -0.51 nm $\pm$ 0.09 nm (mean bias $\pm$ std). Compared to the pixel size of 100 nm, the biases along the x and y-axis can be considered negligible. The observed precisions in these experiments were 4.34 nm $\pm$ 0.08 nm (mean precision $\pm$ std) along the x-axis and 4.32 nm $\pm$ 0.09 nm (mean precision $\pm$ std). Zhang *et al* [76] claimed that smNet could perform axial localizations with a negligible bias and a precision matching the theoretical limit when using the diffraction model. These results show that smNet trained with the accurate vector model matched the performance (negligible bias and precision which is about 1.1 times higher than the theoretical limit along both the axes) of smNet trained using the diffraction model when performing axial localization when the signal and background count were constant.

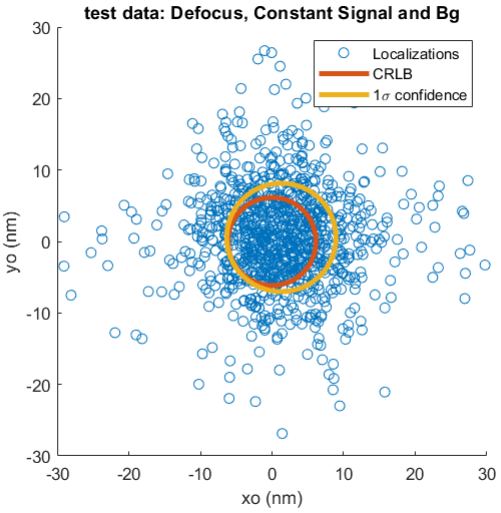


Figure 6.2: Scatter plot of lateral localizations of single emitter molecules with defocus when the signal photon and background count is constant in the presence of constant level of astigmatism.

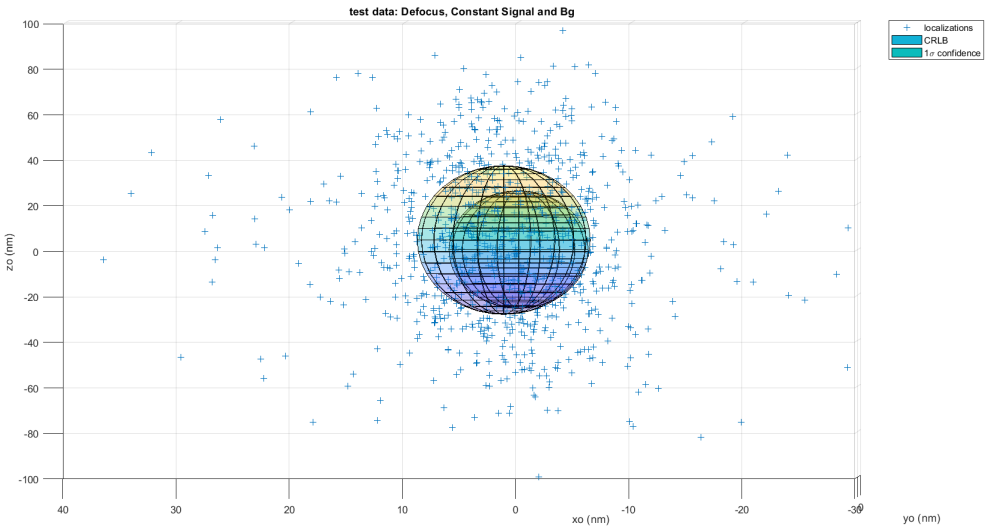


Figure 6.3: Scatter plot of 3D localizations of single emitter molecules when the signal photon and background count is constant in the presence of constant level of astigmatism.

The next experiment was done to characterize the performance of smNet in doing 3D localization when

the signal photon and background count were constant. Figure 6.2 shows the performance of smNet in performing 1000 localization along the x-y axis in the presence of defocus and constant vertical astigmatic wavefront aberration. Figure 6.3 shows the performance of smNet in localizing the single molecules axially and laterally. The inner ellipsoid represents the CRLB (6.24 nm along both the x and y-axis and 25.71 nm along the z-axis) along all the three-axis and the outer ellipsoid represent the precision of smNet. When smNet was used to perform 3D localization on simulated images generated with the diffraction model, Zhang *et al* [76] claimed that the observed bias along the x and y-axis was negligible and bias of $\sim$ 10 nm was observed along the z-axis. It was also claimed that the precision was close to the theoretical limit along all the 3-axes. While performing 3D localization with a smNet model trained with the vector model, it was observed that the biases (mean and std of 10 bias where each bias was calculated from 1000 localizations) were 1.27 nm $\pm$ 0.2 nm along the x-axis, 0.37 nm $\pm$ 0.15 nm along the y-axis and 3.81 nm $\pm$ 0.81 nm along the z-axis. This slight improvement in the bias observed along the z-axis of smNet compared to the diffraction model could be attributed to the better representation of PSF structure and hence a more accurate mapping of the training data to the training labels. The observed precisions were 7.69 nm $\pm$ 0.25 nm along the x-axis, 7.71 nm $\pm$ 0.24 nm along the y-axis and 32.26 nm $\pm$ 0.85 nm along the z-axis. Using the vector model to simulate the training data results in a better performance (bias reduction along the z-axis and precision about 1.2 times higher than the theoretical limit along all the 3-axes) compared to the smNet trained with the diffraction model.

Once the performance of smNet trained with the vector model was characterized in doing 3D localization with a constant signal photon and background count, the next step was to characterize the performance of smNet in doing 3D localization as a function of varying signal photon and background. This characterization was essential as it would give an insight into the performance of smNet under varying SNR conditions. A smNet model was trained with simulated images where each image was assigned a random signal photon count and background which were drawn from 2 Gaussian distributions respectively (details in Chapter 5). Figure 6.4 shows the precision and biases obtained along each of the axes for various levels of signal photon and background count respectively. In characterizing the performance of smNet in doing 3D localization when the signal photon levels were varying, it was observed that the precision obtained closely followed the theoretical limit across all photon levels along the x and y-axis (about 1.35 times the theoretical limit along the x-axis and about 1.25 times the theoretical limit along the y-axis). Along the z-axis, it was observed that the best performance in precision was obtained around the signal photon count of 4000 which corresponded to the mean of the Gaussian curve from which the signal photon count was sampled. This difference in the performance along the x and y-axis and the z-axis arises from the different nature of the two tasks. To perform localization along the x and y-axis, smNet tries to find the position of highest intensity which is the intersection of the all the gradients which corresponds to the centre of ellipse while performing localization along the z-axis is done by creating a correspondence between the PSF shape and the axial position so it requires more training data to accurately map the correspondence between PSF shape and axial position. It was also observed that the bias decreased with increasing photon count for localization along all the axes. This effect shows the limitation of smNet in extracting structural information when the SNR is low. From this, it can be concluded that the width of Gaussian distribution from which signal count was sampled needs to wider so that precise localization along the z-axis can be ensured across all signal photon level. The other alternative is to generate more training samples which will ensure more training samples from the region of very low and very high photon count which would, in turn, ensure more precise localizations at these photon count levels. This method would make the training process computationally more expensive. For the varying level of background count, it is seen that performance degrades faster with a small increase in background as compared to degradation of performance when the the signal photon count is low. This helps us understand that smNet is more robust in dealing with a low level of signal photon count than a high level of background.

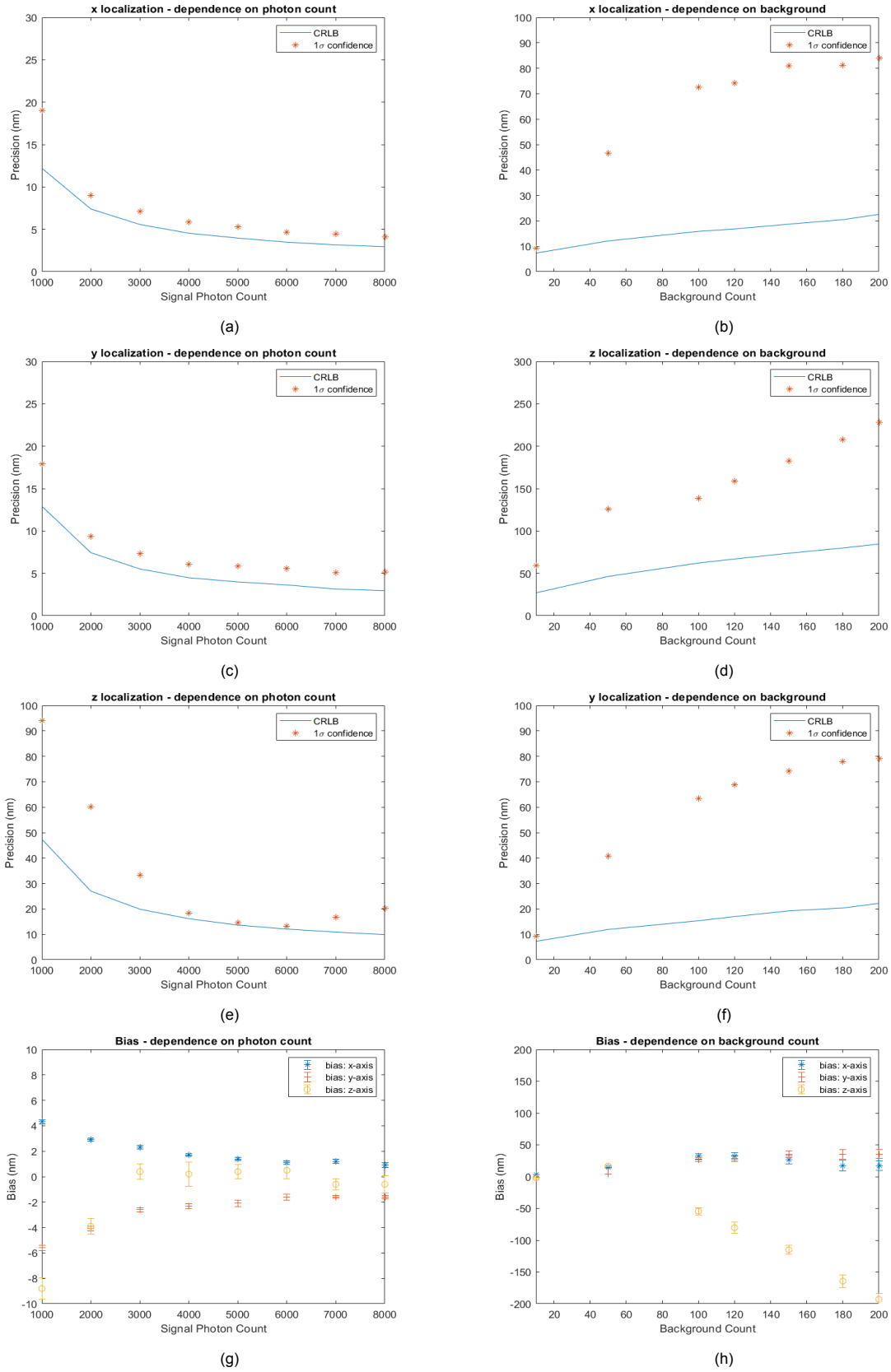


Figure 6.4: (a) Precision of smNet as a function of signal photon count along the x-axis. (b) Precision of smNet as a function of background count along the x-axis. (c) Precision of smNet as a function of signal photon count along the y-axis. (d) Precision of smNet as a function of background count along the y-axis. (e) Precision of smNet as a function of signal photon count along the z-axis. (f) Precision of smNet as a function of background count along the z-axis. (g) Bias of smNet as a function of signal photon count. (h) Bias of smNet as a function of background count. The mean and the std of the biases are calculated from 10 bias and each bias was calculated from 1000 localizations.

Once the performance characterization of smNet in doing 3D localization with variable signal photon count and background was done, the next idea was to characterize the performance of smNet in doing 3D localization while dealing with variable aberrations. In this experiment, the idea was to test the performance of smNet in doing 3D localization in the presence of aberration intensity which was different than the aberration intensity present in the training data. To do this, 2 smNet models were trained. The first smNet model was trained with simulated images having vertical astigmatism intensity of 36 mλ and the second model was trained with images having vertical astigmatism intensity of 72 mλ. Both the models were then tested on the 2 separate datasets. The first dataset comprised of the images with 36 mλ of vertical astigmatism and the second dataset had images with 72 mλ of astigmatism. The results of this experiment are presented in figure 6.5. Both the models are perfectly capable of performing localization along the x and y-axis with precision close to the theoretical limit on the two different dataset. It is also seen that the model trained on images having 36 mλ vertical astigmatism is only capable of precise localization along the z-axis on the dataset having the same aberration intensity. Similarly, the model trained on 72 mλ of vertical aberration does precise localization along the z-axis on a dataset having the same level of vertical aberration. The reason can be attributed to the way smNet learns to do lateral and axial localization. A probable explanation of the way smNet performs lateral localization could be that smNet learns to identify the intersection of the gradients which corresponds to the centre of the elliptical PSF. Irrespective of the ellipticity (ratio of the width of an ellipse to the height of an ellipse) of the PSF (PSFs generated using an optical system having 72 mλ of vertical astigmatic aberration is more elliptical than PSFs generated using an optical system having 36 mλ of vertical astigmatic aberration), the centre of the PSF in an elliptical PSF is the point of intersection of the gradients. So assuming, smNet learns to do lateral localization by identifying the intersection of gradients in a PSF, a smNet model trained on images with 36 mλ of vertical astigmatic aberration is perfectly capable of doing lateral localization of dataset with 72 mλ of aberration as the centre of PSF in both the cases is the intersection of gradients. The reason behind the failure of smNet in doing axial localization on a dataset with different vertical astigmatic aberration is that axial position is learnt by smNet by associating ellipticity to the axial position. So, a PSF generated using an optical system with 72 mλ of vertical aberration has a higher ellipticity than a PSF generated using an optical system with 36 mλ of vertical aberration at the same axial position. So, the ellipticity-axial correspondence learnt by smNet trained with 36 mλ of aberration doesn't hold for smNet trained with 72 mλ of aberration. The bias plot shows that bias decreases with an increase in signal photon count as at lower photon count sufficient information about the elliptical PSF's structure is not available and hence it leads to inaccuracies in predictions.

The next experiment was designed to characterize the performance of smNet in performing 3D localization when the training set had different aberration modes present than the test set. This was done by training the one smNet model with the vertical astigmatic aberration having an intensity of 72 mλ and the second smNet model was trained with a root-mean-squared (R.M.S) aberration intensity of 72 mλ comprising of 5 aberration modes (Noll's index 5,6,7,8,11). Both the smNet models were tested on 2 datasets, the first dataset having only one mode of aberration and the other having 5 modes of aberration of the same RMS intensity of 72 mλ. Figure 6.6 show the performance of both the smNet models on the 2 datasets. The model trained on one aberration mode doesn't perform well on the dataset having 5 modes of aberrations and vice-versa for localization along the x and y-axis. This is because one of the smNet models is trained to find out the centre of the elliptical PSF and it cannot generalize to find the centre of the arbitrary shaped PSF (5 aberration mode) and the second model trained to find the centre of arbitrary shapes isn't well trained to find the centre of the elliptical PSF. For localization along the z-axis, the observation is similar. The ellipticity-axial correspondence learned by smNet doesn't hold when the model is used on a dataset having different aberrations modes than the data used to train the model. So, a model trained on vertical astigmatic aberration does not perform well on the dataset having 5 aberrations modes and vice versa.

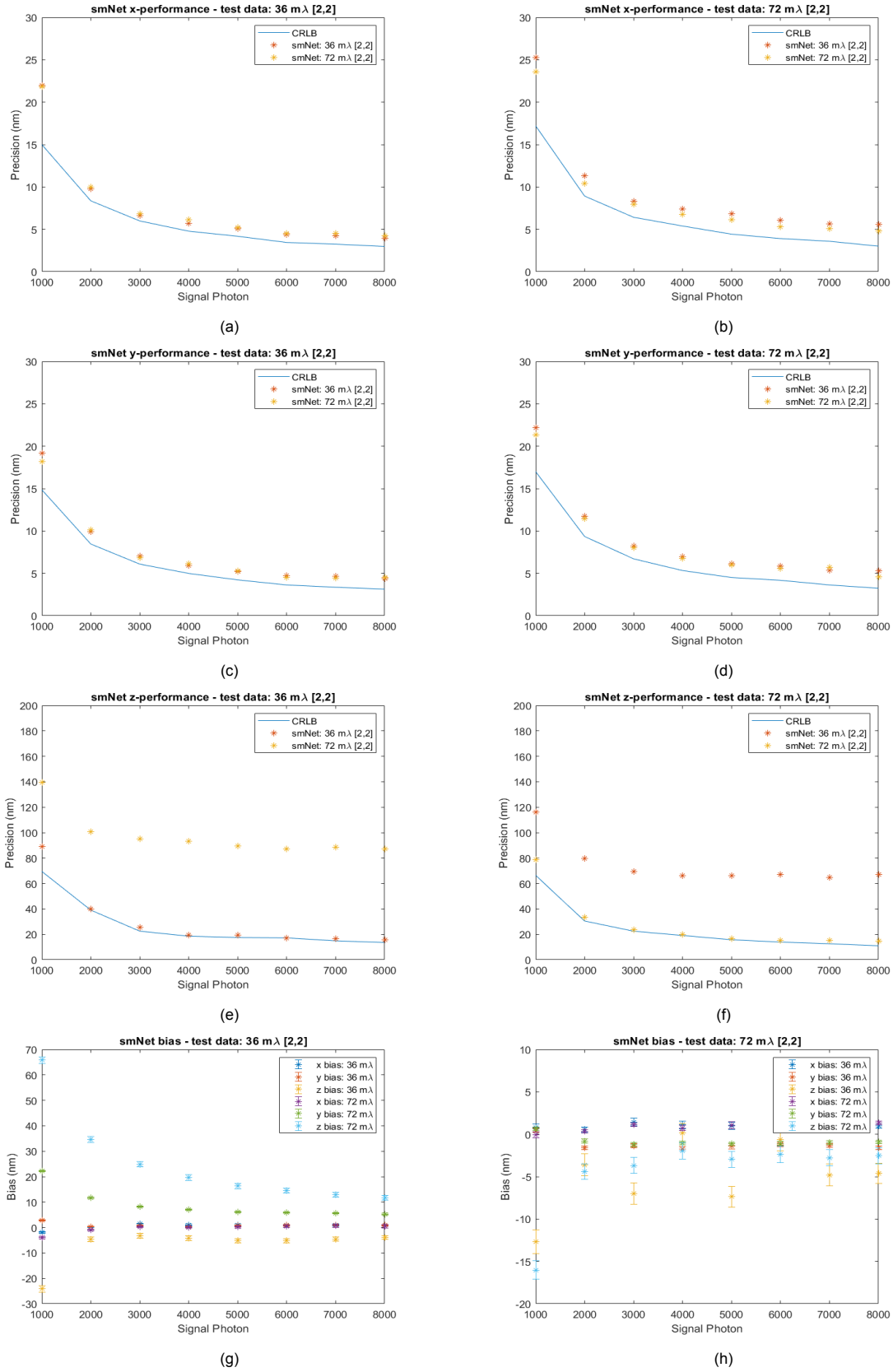


Figure 6.5: (a) Precision of smNet on test data with aberration intensity of 36 $m\lambda$ along the x-axis. (b) Precision of smNet on test data with aberration intensity of 72 $m\lambda$ along the x-axis. (c) Precision of smNet on test data with aberration intensity of 36 $m\lambda$ along the y-axis. (d) Precision of smNet on test data with aberration intensity of 72 $m\lambda$ along the y-axis. (e) Precision of smNet on test data with aberration intensity of 36 $m\lambda$ along the z-axis. (f) Precision of smNet on test data with aberration intensity of 72 $m\lambda$ along the z-axis. (g) Bias of smNet on test data with aberration intensity of 36 $m\lambda$. (h) Bias of smNet on test data with aberration intensity of 72 $m\lambda$. The mean and the std of the biases are calculated from 10 bias and each bias was calculated from 1000 localizations.

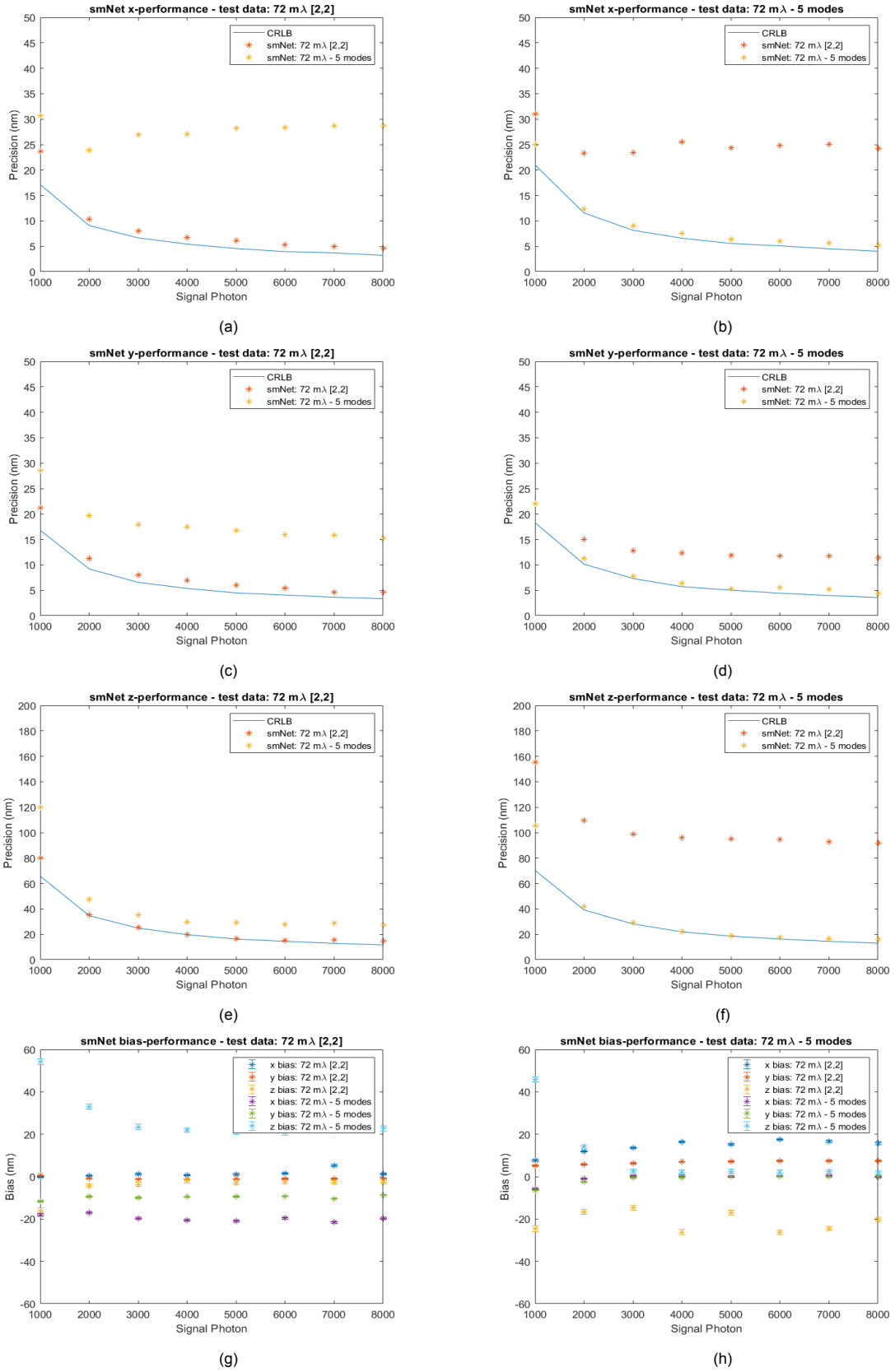


Figure 6.6: (a) Precision of smNet on test data with 1 aberration mode along the x-axis. (b) Precision of smNet on test data with 5 aberration mode along the x-axis. (c) Precision of smNet on test data with 1 aberration mode along the y-axis. (d) Precision of smNet on test data with 5 aberration mode along the y-axis. (e) Precision of smNet on test data with 1 aberration mode along the z-axis. (f) Precision of smNet on test data with 5 aberration mode along the z-axis. (g) Bias of smNet on test data with 1 aberration mode. (h) Bias of smNet on test data with 5 aberration mode. The mean and the std of the biases are calculated from 10 bias and each bias was calculated from 1000 localizations.

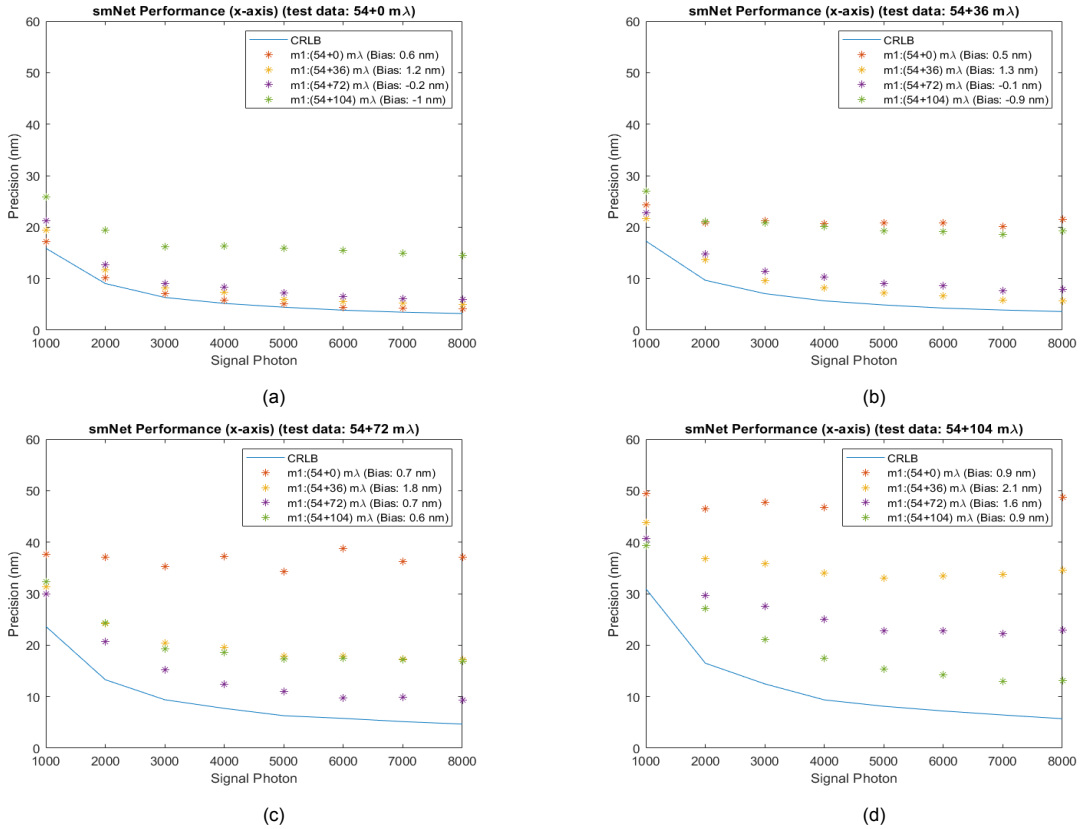


Figure 6.7: Localization performance of smNet along the x-axis for dataset with: (a) vertical astigmatism of 54 mλ + 0 mλ of random aberrations. (b) vertical astigmatism of 54 mλ + 36 mλ of random aberrations. (c) vertical astigmatism of 54 mλ + 72 mλ of random aberrations. (d) vertical astigmatism of 54 mλ + 104 mλ of random aberrations.

Once the characterization of the effect of signal photon count, background, aberration intensity and aberration modes was done, the next idea was to design a method which could ensure robust 3D localization using smNet across the entire parameter space of photon count, background, aberration intensity and modes. The simplest method is to generate a huge amount of training data which covers the entire parameter space. This method is used by Zhang *et al* [76] and this method is extremely compute-intensive and requires a cluster of GPUs to train the network. This makes smNet difficult to use for researchers with limited computational power. This training process uses the brute force approach to train the neural network. According to the brute force approach, the whole parameter space is sampled at once to generate the training data and a lot of training data is required to represent the whole parameter space. Training smNet using such an approach is not optimal and using the conclusions from the previous experiments the idea was to design a better training process which could ensure robust 3D localization across the entire parameter space. The important conclusions from the previous experiments using which the optimal training process was designed are mentioned below:

- To ensure optimal 3D localizations across all signal photon count and background count which are typically seen in an experiment, a broad Gaussian distribution of signal photon and background should be used. This would ensure robust 3D localization in experiments when the signal photon count or background is very low or very high.
- It was also observed for vertical astigmatic aberration mode which is used in experiments to break the symmetry of PSFs in the 2 focal planes, smNet trained on a certain aberration intensity performs robust lateral localization on data having other aberration intensity but axial localization is not precise on a dataset having other aberration intensity.
- It was observed that a smNet model trained with data having one aberration mode cannot perform precise and accurate 3D localization on data having multiple aberrations mode and vice versa.

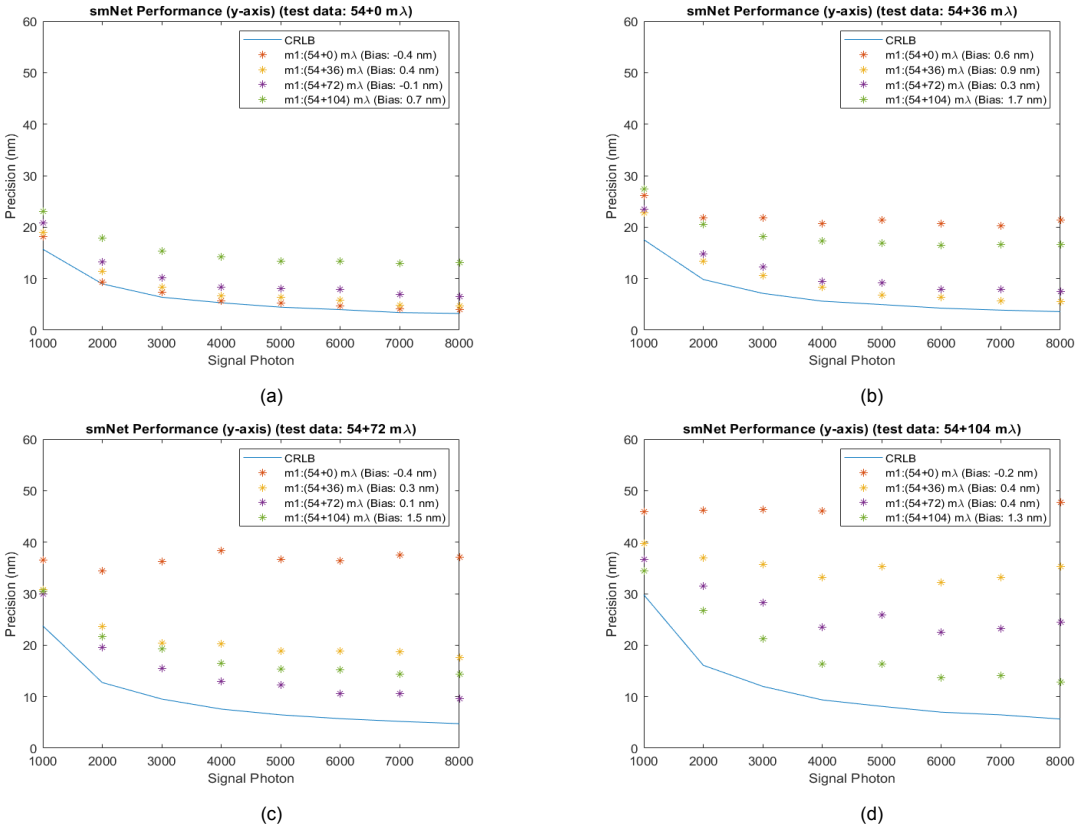


Figure 6.8: Localization performance of smNet along the y-axis for dataset with: (a) vertical astigmatism of 54 mλ + 0 mλ of random aberrations. (b) vertical astigmatism of 54 mλ + 36 mλ of random aberrations. (c) vertical astigmatism of 54 mλ + 72 mλ of random aberrations. (d) vertical astigmatism of 54 mλ + 104 mλ of random aberrations.

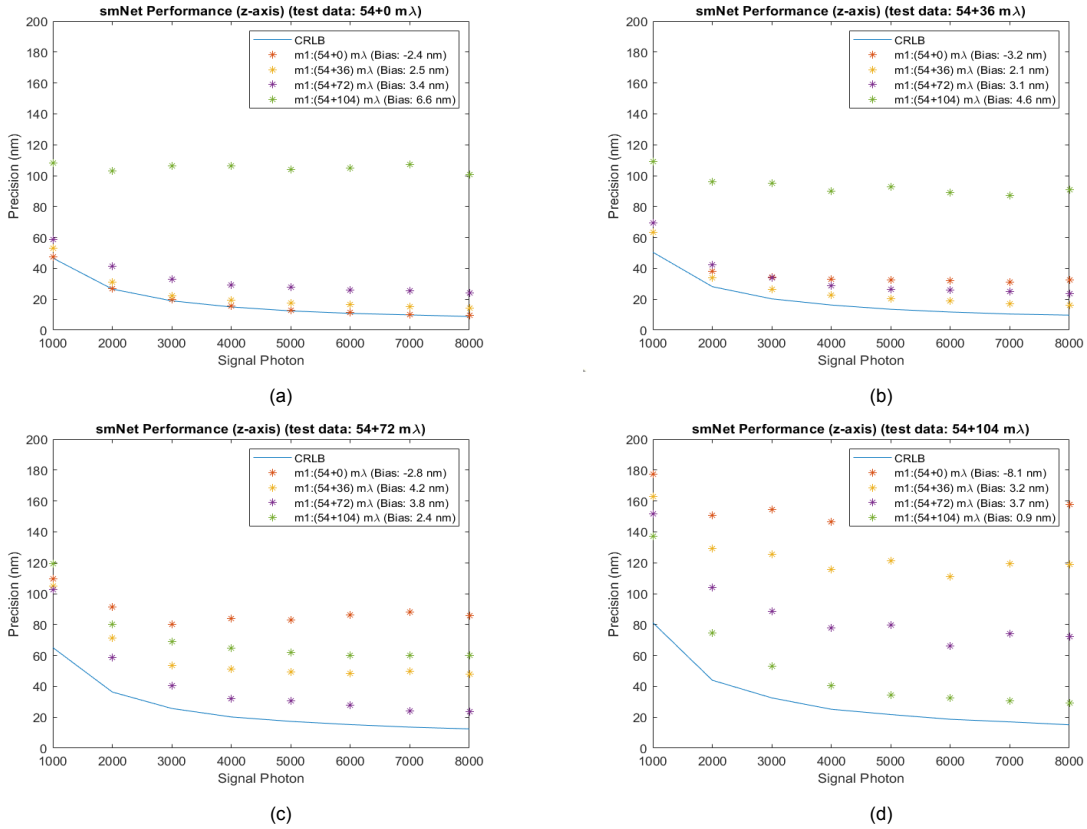
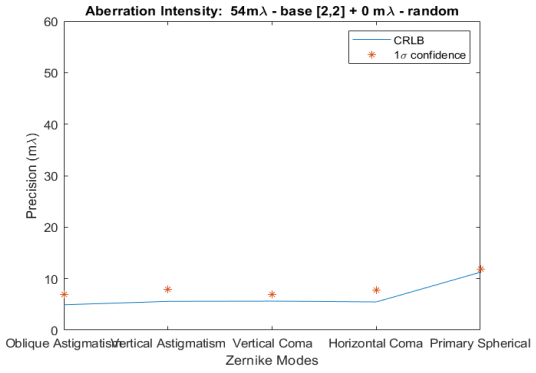


Figure 6.9: Localization performance of smNet along the z-axis for dataset with: (a) vertical astigmatism of 54 mλ + 0 mλ of random aberrations. (b) vertical astigmatism of 54 mλ + 36 mλ of random aberrations. (c) vertical astigmatism of 54 mλ + 72 mλ of random aberrations. (d) vertical astigmatism of 54 mλ + 104 mλ of random aberrations.

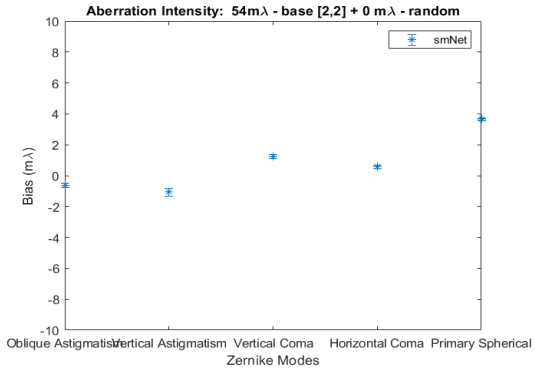
The training process was made efficient by using these conclusions. The idea was to split up the whole parameter space and train multiple models to deal with a section of the parameter space. A pre-estimate (calibration run) of the vertical astigmatic aberration used to break the symmetry was the apriori knowledge that was required which was also done using smNet (discussed in details in subsection 6.1). Figures 6.7, 6.8 and 6.9 represents the performance of smNet in doing 3D localization over a range of aberrations intensity over a range which varies from 0 mλ to 104 mλ which is well over the Marechal's diffraction criterion of 72 mλ using the splitting the parameter space approach. The training of a model was done by generating dataset having a base vertical aberration of the estimated intensity (54 mλ in this case) and adding 5 modes of random aberrations (Noll's index 5,6,7,8,11) on top of it of a certain RMS intensity. Splitting the parameter space in this way ensures robust performance (precision of best model less than 1.7 times the theoretical limit across all aberration intensities) across all aberration intensity level of all aberration modes and the training process isn't as computationally expensive as the brute force method. To cover the whole parameter space, Zhang *et al* [76] needed 1 million images while training one model using this technique requires only 160,000 images and since 4 such models are used to cover the entire parameter space the total images required are 640,000. Since we train each model separately, each model is trained with only about one-tenth of the images needed to train the original smNet model.

Characterization of aberration estimation performance

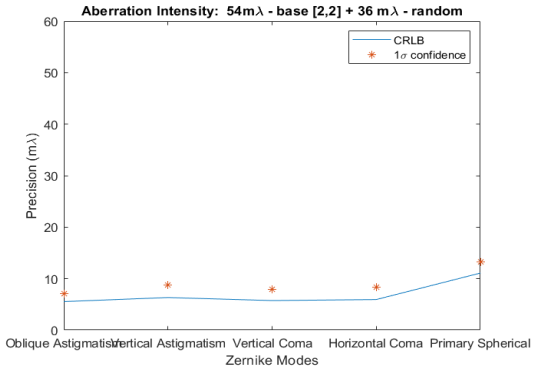
Once the smNet's ability to perform 3D localization over a vast range of physical conditions was characterized, the next step was to characterize the performance of smNet in doing aberration estimation. This step was essential as the proposed pipeline which was used to do 3D localization had a model selector module which estimates the aberrations present in the image and then makes a selection of the most suitable smNet model for 3D localization depending on the aberration intensities. Figure 6.10 shows that smNet can be used to perform accurate and precise aberration estimation over a broad range of aberrations intensities. For aberration intensities below the Marechal's limit of 74 mλ it is observed that smNet can perform aberration estimation with precisions close to the theoretical limit with very high accuracy. Even beyond the Marechal's limit smNet can perform aberration estimation with considerable precision and very high accuracy. This characterization was done for oblique astigmatism, vertical astigmatism, horizontal coma, vertical coma and primary spherical aberrations. It was essential for the robust performance of the pipeline that the smNet could perform aberration estimation over a large range of imaging conditions. Figure 6.11 shows the performance of smNet in estimating horizontal astigmatic aberrations over a large range of photon count and background. For the other modes of aberrations, a similar trend was observed. It can be be seen that smNet performs aberrations estimation with precisions close to the theoretical limit for a large range of signal photon count with very high accuracy (highest bias of 2.68 mλ and an average bias of -0.37 mλ). For background, it was seen that smNet could do precise aberration estimation close to the theoretical limit over a large range of background count but when the background was very high ($\sim$ 200) the precision decreased slightly. For the varying level of background, smNet could perform aberration estimation very accurately where the worst performance was a bias of 4 $m\lambda$ for a background level of 200 and the average bias for aberration estimation for was 0.75 $m\lambda$. Figure 6.12 shows that smNet can estimate oblique and vertical astigmatism more accurately when the emitters are not exactly at the focus because of the spherical nature of the PSF spots near focus. For primary spherical aberration, smNet could perform more accurate estimation when the emitters were near focus. This might be happening because the secondary rings associated with spherical aberration might be going out of 16 pixel $\times$ 16 pixel region of interest when the single emitters are away from the focus as the size of the spot becomes bigger and blurred when the emitters are away from the focus.



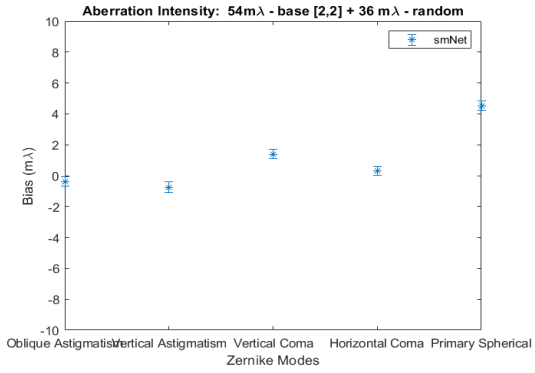
(a)



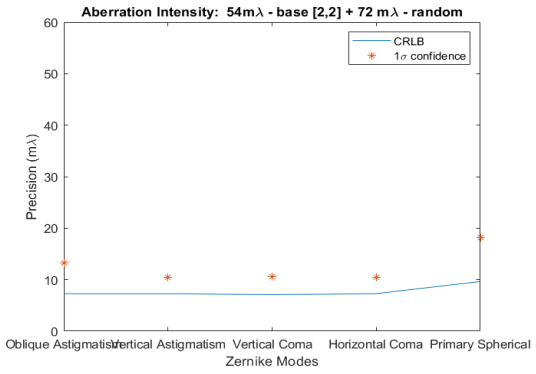
(b)



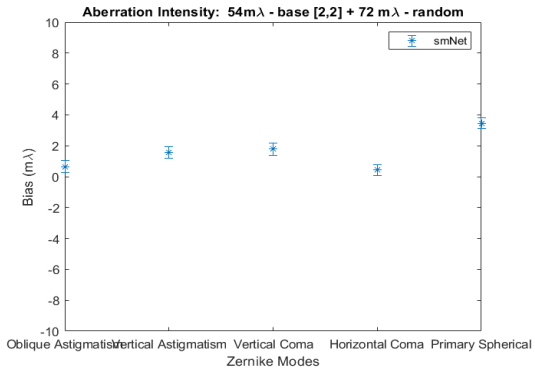
(c)



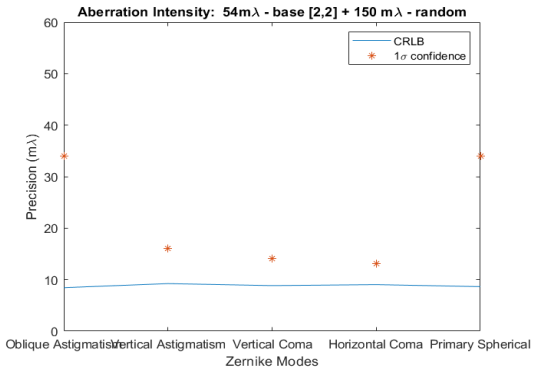
(d)



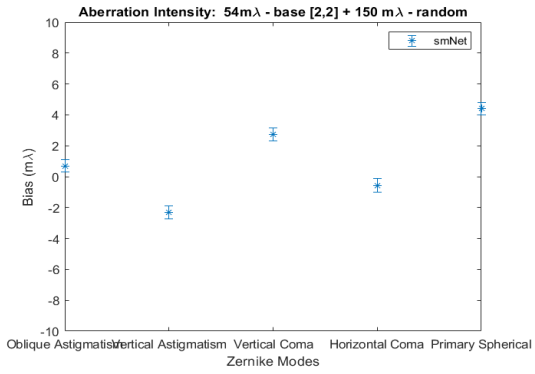
(e)



(f)

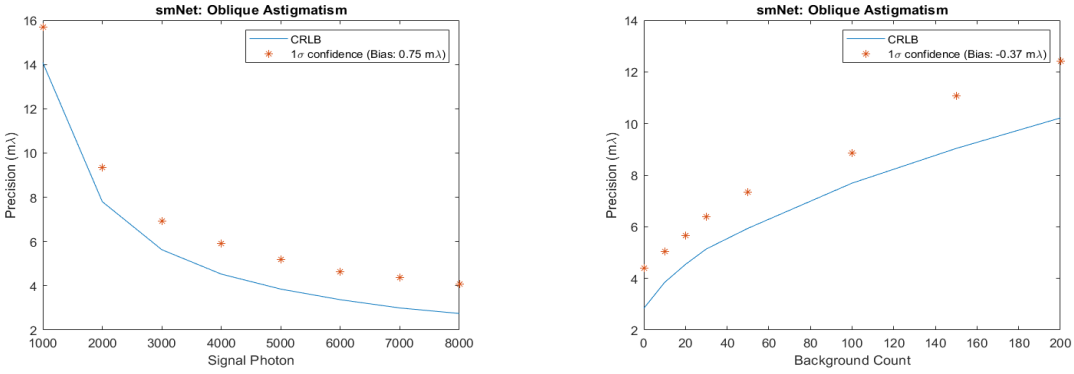


(g)



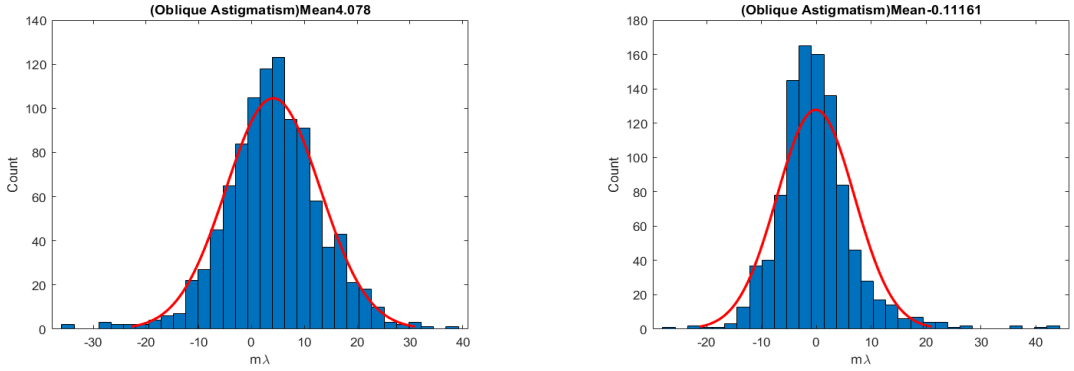
(h)

Figure 6.10: (a) Precision of smNet in estimating aberrations ($W_{RMS} = 0m\lambda$) (b) Bias of smNet in estimating aberrations ($W_{RMS} = 0m\lambda$) (c) Precision of smNet in estimating aberrations ($W_{RMS} = 36m\lambda$) (d) Bias of smNet in estimating aberrations ($W_{RMS} = 36m\lambda$) (e) Precision of smNet in estimating aberrations ($W_{RMS} = 72m\lambda$) (f) Bias of smNet in estimating aberrations ($W_{RMS} = 72m\lambda$) (g) Precision of smNet in estimating aberrations ($W_{RMS} = 150m\lambda$) (h) Bias of smNet in estimating aberrations ($W_{RMS} = 150m\lambda$). The mean and std of biases are computed from 10 bias and each bias is calculated from 1000 estimations.



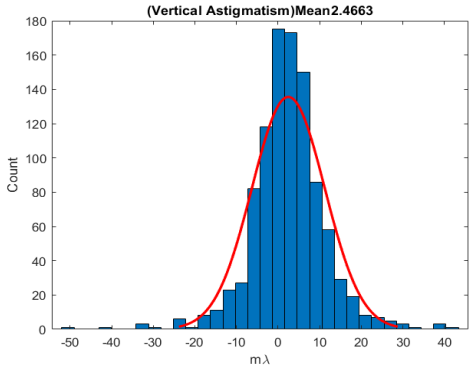
(a) (b)

Figure 6.11: Estimation of oblique astigmatism as a function of (a) signal photon (b) background

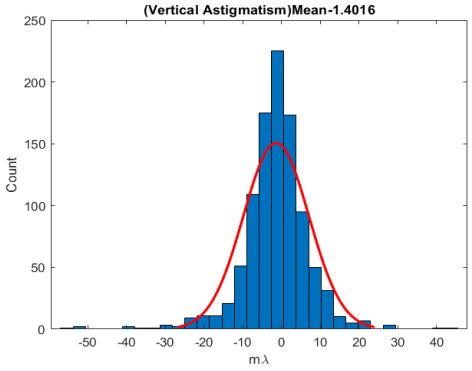


(a) Emitters at focus

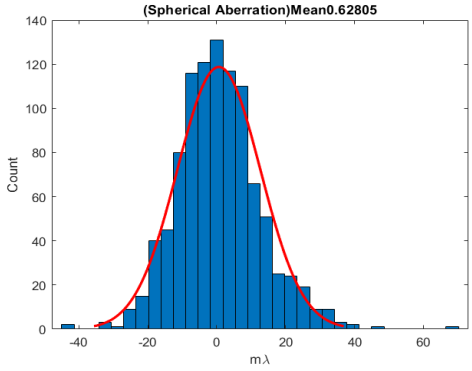
(b) Emitters away from focus



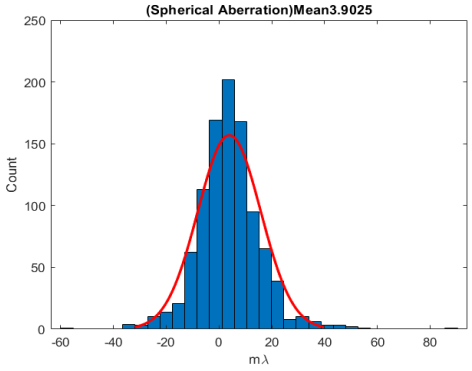
(c) Emitters at focus



(d) Emitters away from focus



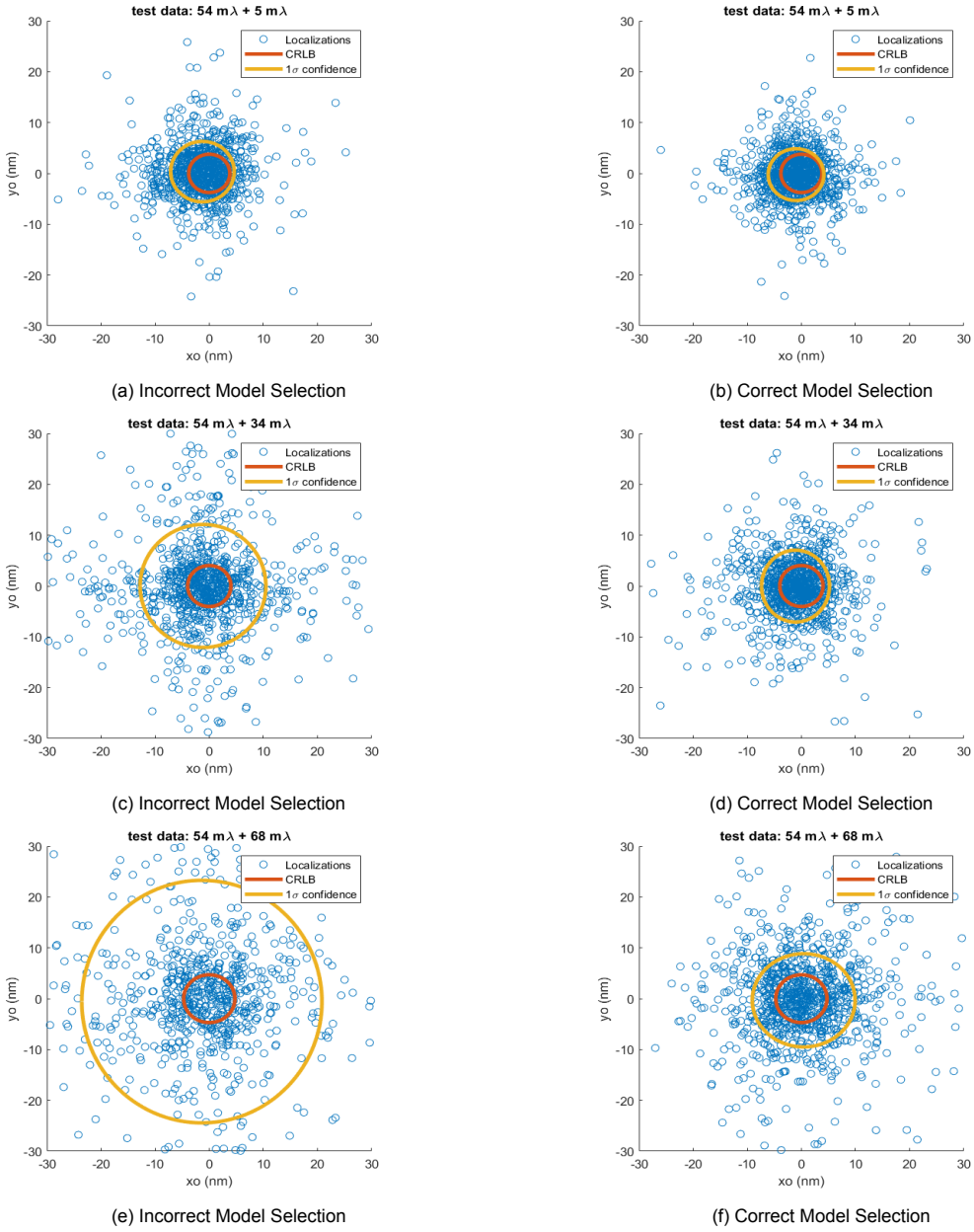
(e) Emitters at focus



(f) Emitters away from focus

Figure 6.12: Aberration estimation at the focus and away from the focus

6.2. Pipeline for Simulator Learning



(a) Incorrect Model Selection

(b) Correct Model Selection

(c) Incorrect Model Selection

(d) Correct Model Selection

(e) Incorrect Model Selection

(f) Correct Model Selection

Figure 6.13: Comparison of the scatter plots of localization as a function of model selection for test data (a,b) 54 mλ vertical aberration + 5 random Zernike modes of W_{rms} = 5 mλ (c,d) 54 mλ vertical aberration + 5 random Zernike modes of W_{rms} = 34 mλ (e,f) 54 mλ vertical aberration + 5 random Zernike modes of W_{rms} = 68 mλ

The main idea was to design a pipeline using which smNet trained on images generated using the accurate vector model could be used to perform robust 3D localization for a vast range of physical conditions. The proposed pipeline (details in Section 5) would take in a 16 pixels $\times$ 16 pixels region of interest containing a single emitter as an input. The model selection module built using the smNet's aberration estimation pipeline would be used to estimate the aberration intensities present in the image of the single emitter. Once the estimation of the aberration intensities is made, selection of the best smNet model to perform 3D localization would be done. The selected smNet model would be used to localize the single emitter axially and laterally. In this section, the performance of the proposed pipeline is tested. The model selection module estimates the aberration intensities of the 5 aberration modes (Noll's index 5,6,7,8,11). Once the estimates are made, a constant vertical astigmatism aberration which was determined in the calibration run is subtracted from the aberration estimate of vertical aberration. Next, the RMS aberration intensity is calculated using the estimates of the 5 aberration modes and then the appropriate model is selected based on the difference of the estimated RMS and the RMS intensities of the trained model. The model which has the smallest difference is selected as the most appropriate model to 3D localization. Table 6.1 shows the performance of the model selection module on the test data set with different RMS aberration intensity levels (54 mλ base vertical astigmatism + random aberration intensities). It is observed that instead of 4 models which each covers 36 mλ of the 108 mλ aberration intensity parameter space if 10 models each covering $\sim$ 10 mλ of the aberration intensity parameter space then the model selection algorithm would be more accurate and hence the 3D localization would be more precise and accurate.

Table 6.1: Performance of the model selection mechanism based on aberration estimation

Aberration	Model 1 (54+0 mλ)	Model 2 (54+36 mλ)	Model 3 (54+72 mλ)	Model 4 (54+108 mλ)
(54+8 mλ)	63.2 %	35.1 %	1.6 %	0.1 %
(54+16 mλ)	37.5 %	61.3 %	1.2 %	0 %
(54+24 mλ)	10.4 %	88 %	1.6 %	0 %
(54+32 mλ)	1.2 %	94.2 %	4.6 %	0 %
(54+40 mλ)	0.5 %	91.3 %	8 %	0.2 %
(54+48 mλ)	0.2 %	80.1 %	19.4 %	0.3 %
(54+56 mλ)	0.1 %	52.6 %	46.6 %	0.7 %
(54+64 mλ)	0 %	26.7 %	71.8 %	1.5 %
(54+72 mλ)	0 %	11.1 %	83.9 %	5 %
(54+80 mλ)	0 %	4.7 %	82 %	13.3 %
(54+88 mλ)	0 %	4.3 %	67.5 %	28.2 %
(54+96 mλ)	0 %	3.4 %	51.2 %	45.4 %
(54+104 mλ)	0 %	2.2 %	42.1 %	55.7 %
(54+112 mλ)	0 %	2.1 %	31.5 %	66.4 %

Table 6.2: Performance of the pipeline in localization along the x-axis

Test Data	Selected Model	Accuracy	CRLB (nm)	1σ (nm)	Bias (nm)
(54+8 mλ)	model 1 (54 mλ + 0 mλ)	65.7 %	6.35	8.90	0.29
(54+34 mλ)	model 2 (54 mλ + 36 mλ)	94.2 %	6.48	8.98	1.13
(54+68 mλ)	model 3 (54 mλ + 72 mλ)	81 %	8.69	14.30	0.26

Table 6.3: Performance of the pipeline in localization along the y-axis

Test Data	Selected Model	Accuracy	CRLB (nm)	1σ (nm)	Bias (nm)
(54+8 mλ)	model 1 (54 mλ + 0 mλ)	65.7 %	6.41	8.48	-0.42
(54+34 mλ)	model 2 (54 mλ + 36 mλ)	94.2 %	6.55	8.91	1.00
(54+68 mλ)	model 3 (54 mλ + 72 mλ)	81 %	8.38	14.79	0.52

Tables 6.2, 6.3 and 6.4 show the performance of the pipeline in performing 3D localization after selecting the appropriate models trained to handle a certain range of aberration intensities. Figure 6.13 visually

Table 6.4: Performance of the pipeline in localization along the z-axis

Test Data	Selected Model	Accuracy	CRLB (nm)	1σ (nm)	Bias (nm)
(54+8 mλ)	model 1 (54 mλ + 0 mλ)	65.7 %	22.19	23.9	-3.7
(54+34 mλ)	model 2 (54 mλ + 36 mλ)	94.2 %	23.10	42.82	2.65
(54+68 mλ)	model 3 (54 mλ + 72 mλ)	81 %	25.66	65.36	3.81

show the effectiveness of the designed pipeline and selection of the correctly trained model improved the precision and accuracy with which the correct smNet model could perform localization compared to the incorrect model. Hence, it was verified that the designed pipeline not only reduced the training time, computational load but also it could be used to robust 3D localizations across a vast parameter space of random signal photon count, background, aberration modes and aberration intensities.

6.3. Simulator Learning

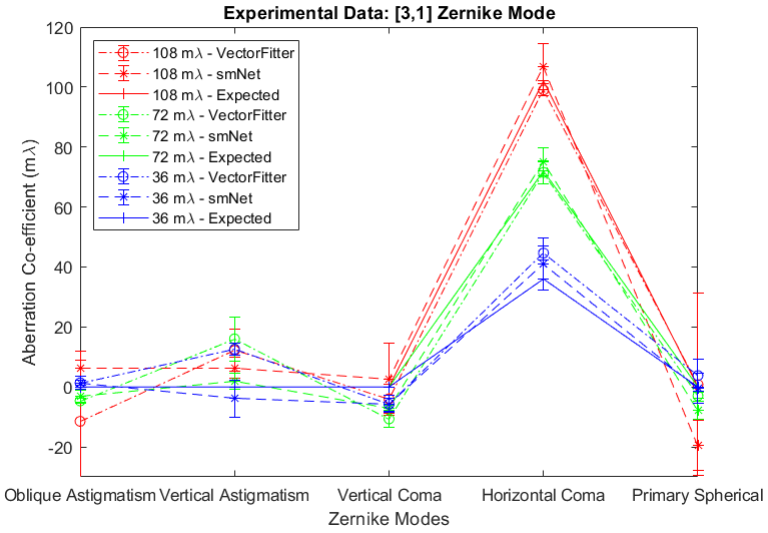


Figure 6.14: Comparison of the performance of smNet and vector fitter in estimation of vertical coma in experimental data.

Zhang *et al* [76] used interpolation-based data augmentation on experimental data to train the neural networks to make predictions on experimental data and a simplified diffraction model and a phase retrieval method to extract the pupil function to train smNet on simulated data to make predictions on simulated and experimental data. The concept of simulator learning – training the neural network with purely simulated data without using any apriori knowledge from the experimental data (done using the vector model) and testing on experimental data was tested using smNet. The performance of smNet was compared to a vector fitter algorithm [67] in estimating the aberrations present in the experimental data. The first experiment was done where the performance of smNet trained on simulated data was compared to the vector fitter algorithm in estimating vertical coma. An SLM was used in the experimental setup to convert uncorrected images with arbitrary aberrations to images of a single emitter having a certain aberration intensity of certain aberration mode. In the first experiment, using SLM only vertical coma was present and the rest of the aberration modes were set to zero and the results are shown in figure 6.14. In the next experiment, the SLM was used to generate experimental data having only vertical astigmatism and the rest of the aberration modes set to zero. The performance of smNet and the vector fitter on this experimental data is shown in the figure 6.15. In the third experiment, experimental data with only spherical aberration was generated and the smNet and vector fitter algorithm was used to estimate the aberrations present in the images and the result is shown in figure 6.16. It was observed that for the first two experiments both smNet and the vector fitter algorithm estimated the aberration modes and the aberration intensities quite accurately. In both the experiments, it was seen that smNet performs slightly better than the vector fitter algorithm. In the third experiment, both smNet

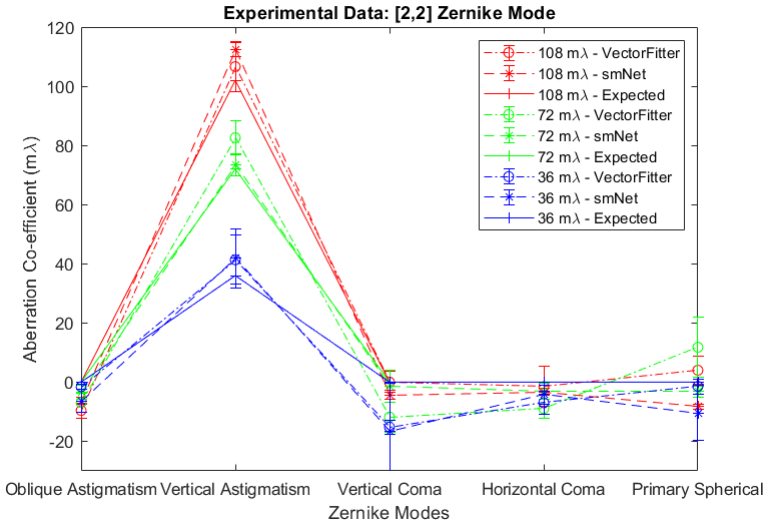


Figure 6.15: Comparison of the performance of smNet and vector fitter in estimation of vertical astigmatism in experimental data.

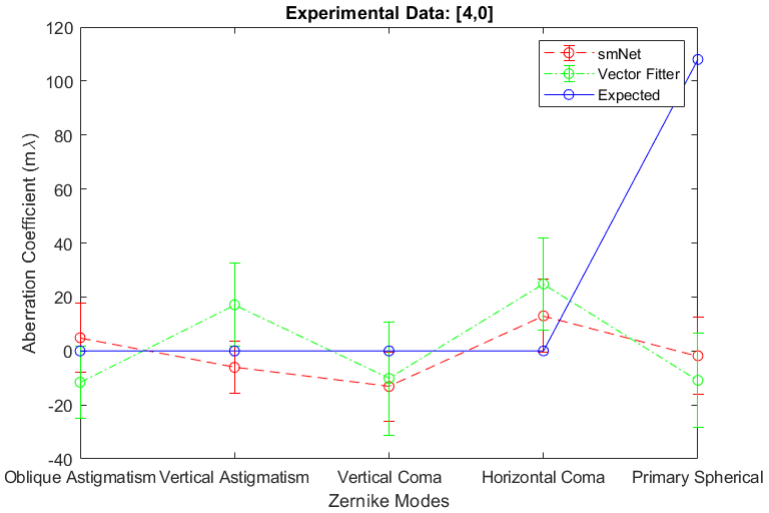


Figure 6.16: Comparison of the performance of smNet and vector fitter in aberration estimation in experimental data.

and the vector fitter algorithm fails to estimate the correct aberration mode and the correct aberration intensities shown in figure 6.16. It is seen in figure 6.19 that the rings which are associated with primary spherical aberration are not contained with the bound of the 16 pixels $\times$ 16 pixels. Hence, the features associated with the image of a single emitter corrupted with primary spherical aberrations is missed by both smNet and vector fitter algorithm and both the methods produce wrong estimates. The final experiment was done on uncorrected images and figure 6.17 shows the performance of both smNet and the vector fitter algorithm in estimating aberrations on uncorrected images. It was observed that smNet and vector fitter algorithm produced estimates which were different. Figure 6.18 shows that there was a presence of higher aberration modes in the uncorrected images. This was the reason for mismatch in the estimates produced by smNet and the vector fitter algorithm as the smNet model was only trained with only 5 aberration modes (Noll's index 5,6,7,8,11). The speed of both algorithms was also compared. It took smNet 0.02 secs to estimate the aberrations present in the experimentally obtained images while it took the vector fitter algorithm 44.69 secs to do the same task. smNet is significantly much faster than the vector fitter algorithm. Another benefit of using smNet was that smNet could estimate the aberrations present in the image using a single image while the vector fitter algorithm needed a stack of images taken at different axial positions to do the same. In these experiments,

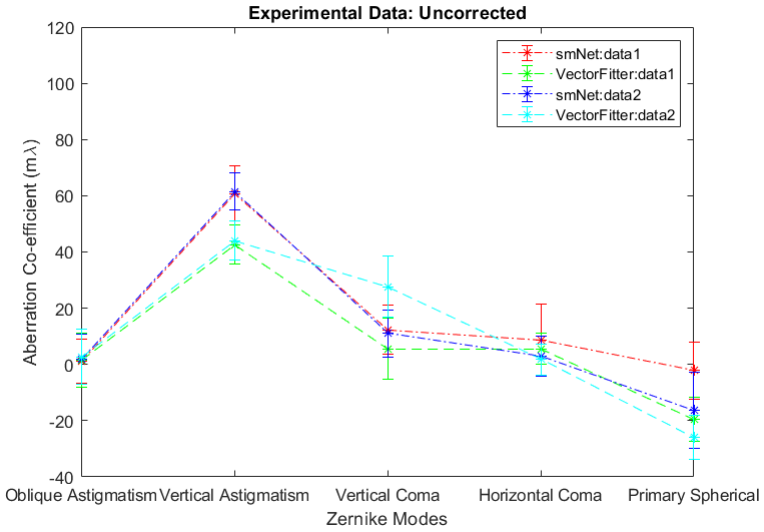


Figure 6.17: Comparison of the performance of smNet and vector fitter in aberration estimation in experimental data.

it was shown that smNet which was trained on simulated data could be used directly on experimental data without any retraining or domain adaptation [30].

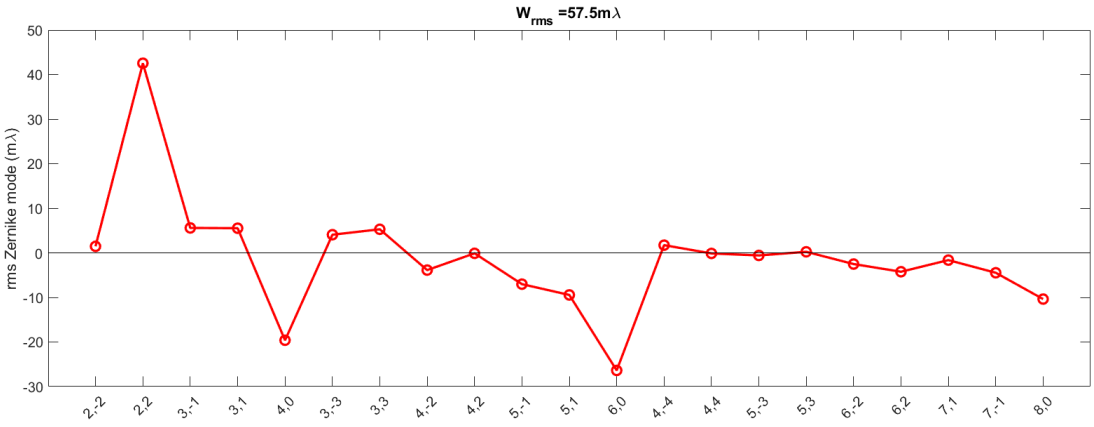
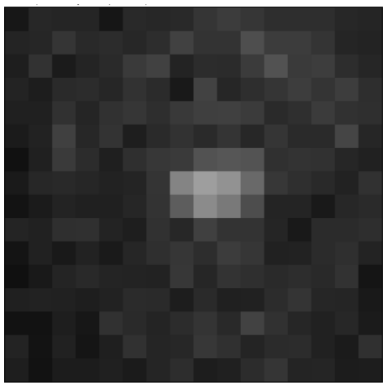
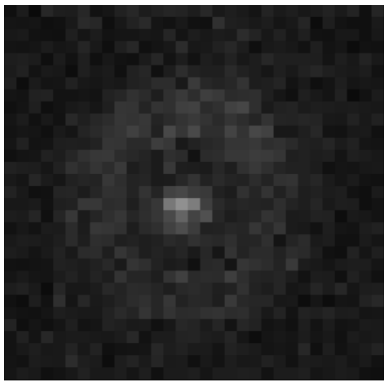


Figure 6.18: Evidence of presence of higher aberration mode in uncorrected data which causes the mismatch of the estimate of smNet and vector fitter algorithm.



(a) Image with horizontal field of view of 16 pixels.



(b) Image with horizontal field of view of 31 pixels.

Figure 6.19: Comparison of the image of single emitter with primary spherical aberration with the horizontal field of view of 16 and 31 pixels respectively.

7

Conclusions and Recommendations

7.1. Conclusions

The aim of the thesis was to find out if smNet trained with the simulated images generated using the accurate vector model could perform accurate and precise 3D localization for a wide range of parameters and to design a pipeline using which smNet trained on simulated images could be used to be used on experimental data without any retraining.

The time taken to generate a super-resolution image is limited by the speed of the algorithm which localizes the emitters (discussed in detail in chapter 1 and chapter 2). Fluorescent microscopes are used to generates frames of sparsely distributed fluorescent protein. The centres of the fluorescent proteins are then localized using localization algorithms and with thousand of such frames and information about the centre of each of the protein label a synthetic super-resolved image is created which reveals information about the finer biological structures which could not be seen before using light microscopy as the resolution was limited by Abbe's diffraction limit. The image formation process in a fluorescent microscope is discussed in detail in chapter 3. There is a mathematical limit on how precise an unbiased estimator can be which is called the Cramer-Rao lower bound and this concept is also discussed in chapter 3.

Since the speed of generation of super-resolved images is constrained by the speed of the localization process, smNet was developed which claimed to improve the speed of localization while performing accurate and precise localization. This would improve the speed of generation of a super-resolution image without impacting the quality of the synthetically created image. smNet was a deep learning method which was designed to tackle multiple problems in the field of localization microscopy. It could perform 3D localization, estimation of the orientation angle of a dipole and estimation of the optical aberrations present in a system. This was done by multiplexing each of the problems and making the complexity additive instead of multiplicative. The design of the smNet neural network and the working and purpose of all it's building blocks are discussed in details in chapter 4

smNet was trained on either augmented experimental data or simulated data which was generated using an erroneously simple simulation model. The problem with using augmented experimental data was that an interpolation-based augmentation technique was used which would not provide an accurate representation of the entire parameter space. Generation of simulated images for training using the simple diffraction model also had the same shortcoming of not representing the reality with accuracy and needed apriori information about the pupil function from experimental data. Training smNet with an accurate vector model would ensure smNet learnt the correct intricate details and in principle should be more robust in doing localization and it would not require any apriori information from the experimental data. The performance of smNet, trained with the vector model generated simulation images, was characterized for varying physical conditions and the development of a pipeline which made the training process more efficient while maintaining the accuracy and precision of localization and the testing of the idea of simulator-learning was discussed in chapter 6. It was found out that the pipeline which was designed could be used to perform 3D localization accurately and precisely for a wide range of

parameters. The idea of splitting the parameter space and using multiple smNet models to deal with a section of the parameter space made the training process efficient and reduced the computational load and training time.

7.2. Recommendations

All the experiments were done with only 5 aberration modes to show as a proof of concept that smNet can be trained using the vector model generated images to do precise and accurate 3D localization. smNet can be used to estimate up to 21 aberration modes, so the same concept of splitting the parameter space and proposed pipeline can be extended when images are corrupted with many more aberration modes. The next step would be to test the working of the pipeline when smNet is trained with more aberration modes and to find the effect of the scaling up of the parameter space with the amount of training data that is required.

The localization pipeline which was proposed was used independently to generate a localization list. An important step forward would be to integrate the proposed pipeline into the localization microscopy pipeline and use it end to end on images of biological samples to generate a super-resolution image. The quality of the generated super-resolution image would help us understand the performance of smNet when it is a part of the localization pipeline.

Another important step would be to do the experiments with bigger ROIs (32 pixels $\times$ 32 pixels) to see if the accuracy with which primary spherical aberrations are estimated improves. In all of these experiments, the ROI was kept small to ensure that the only one single emitter was present per ROI. It would be interesting to see how the presence of the trailing edges of other single emitters in the ROI would affect the estimation precision and accuracy. This would help researchers find out how effective smNet can be in doing estimation in a dataset where the labelling density is very high. Another interesting step forward would be to characterize how effective smNet is in handling multi-emitter localization and to find out what modification needs to done to make smNet perform multi-emitter localization.

smNet was also designed to perform estimation of the orientation angle of a dipole emitter. Training smNet with the accurate vector model for orientation estimation and its performance characterization is also something which needs to be explored. In this thesis, only smNet's ability to perform 3D localization and aberration estimation has been studied in details.

Another recommendation would be to make the smNet more accessible to researchers with limited hardware resources. One method of doing it would be by application of model compression techniques [20], [34], [38] . Model compression techniques such as pruning and weight sharing have been successfully implemented in object detection problems which involve both classification and regression. 3D localization and aberration estimation largely are regression problems and hence the same model compression techniques can be tested on smNet. Applying these model compression techniques would make the total number of learnable parameters even smaller thereby making the training process of smNet computationally much cheaper. The model compression techniques would also make the final model size smaller making it possible to implement smNet as an edge computing solution where the model can be deployed offline integrating it with the microscopes data acquisition software.

8

Appendix A

Training and Test Data Representation

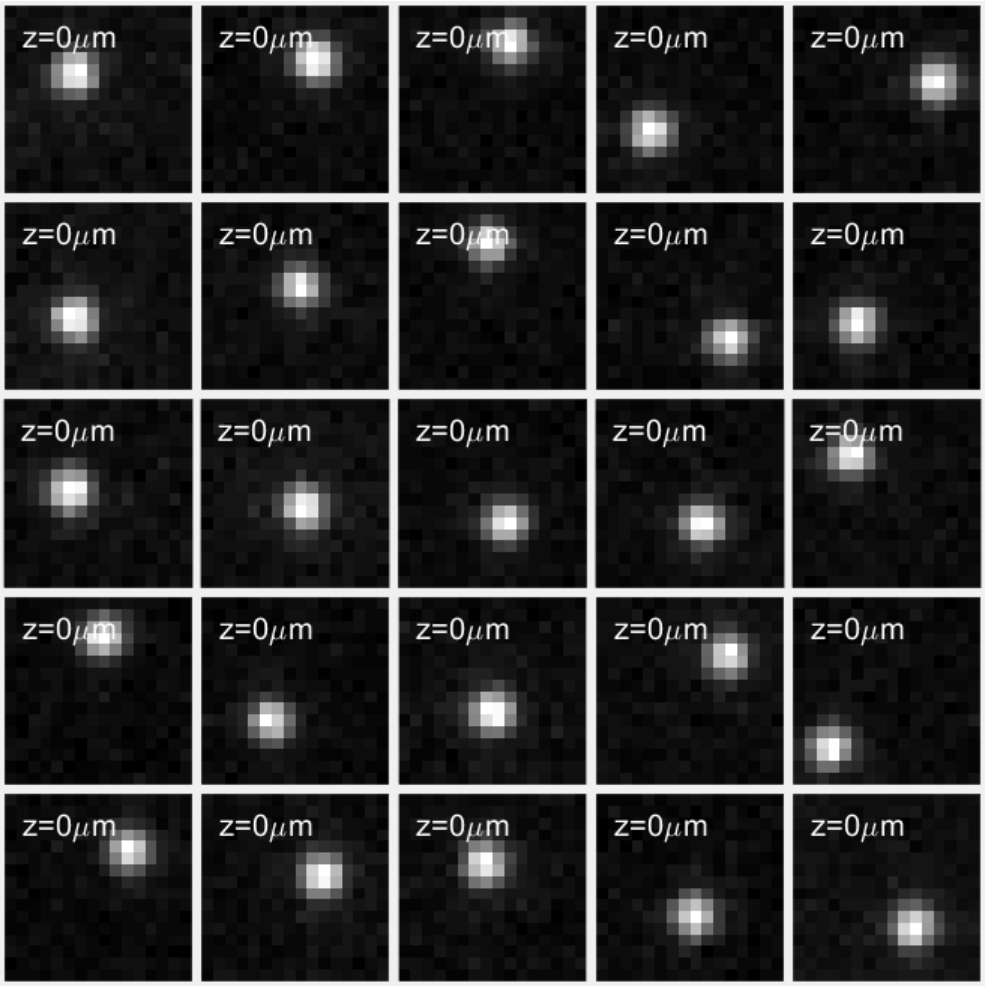


Figure 8.1: Training and Test Data Representation: Experiment 1 - Characterizing the performance of smNet in performing lateral localization when emitters are at the focus and the background and signal count are constant

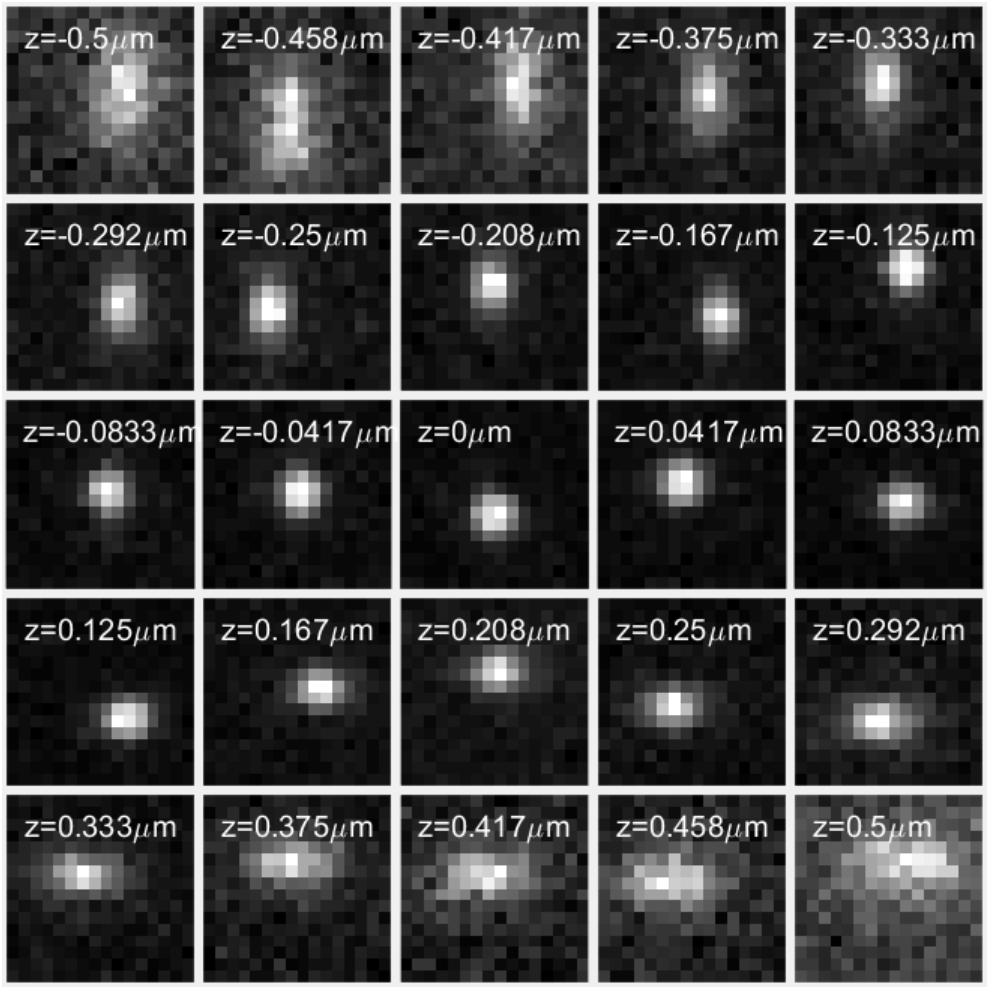
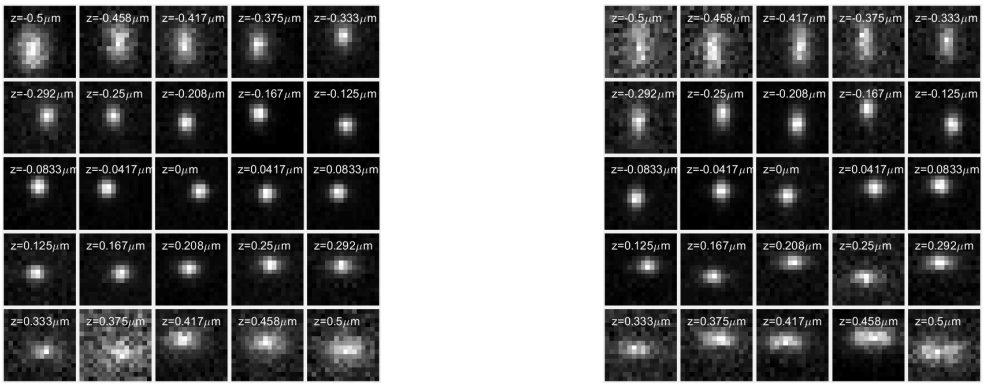
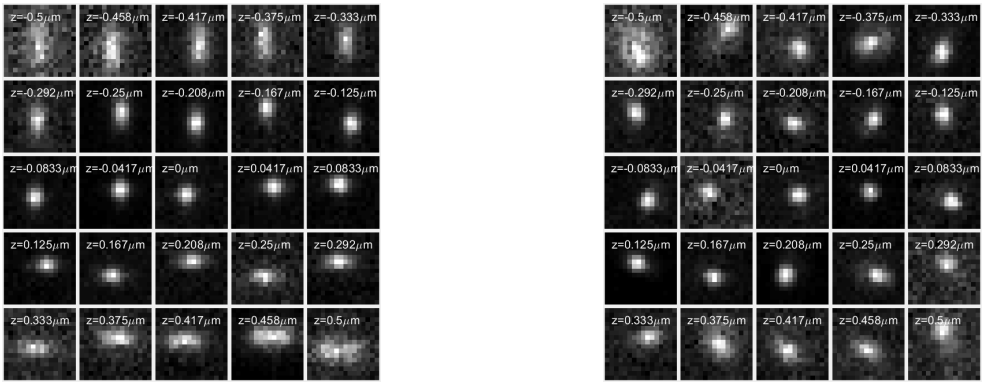


Figure 8.2: Training and Test Data Representation: Experiment 2 - Characterizing the performance of smNet in performing 3D localization when emitters are at the focus and the background and signal count are constant



(a) (b)

Figure 8.3: Training and Test Data Representation (a) with 36 mλ of vertical astigmatic aberration (b) with 72 mλ of vertical astigmatic aberration



(a) (b)

Figure 8.4: Training and Test Data Representation (a) with 72 mλ of vertical astigmatic aberration (b) with 72 mλ of random aberrations (Noll's Index 5,6,7,8,11)

Training Curve

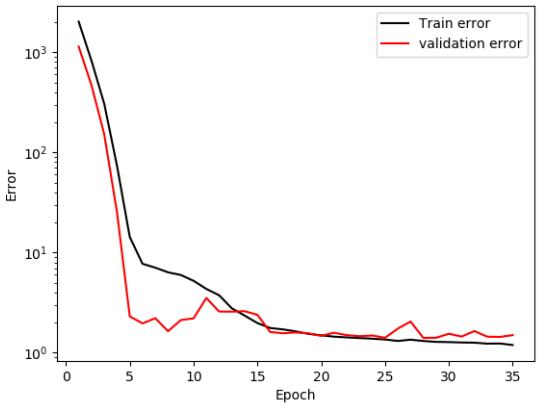
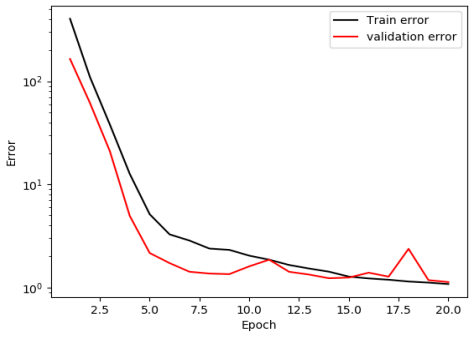
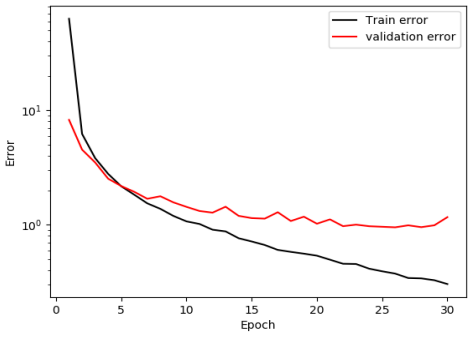


Figure 8.5: Training Curve : Experiment 1 - Lateral localization when signal photon and background are constant

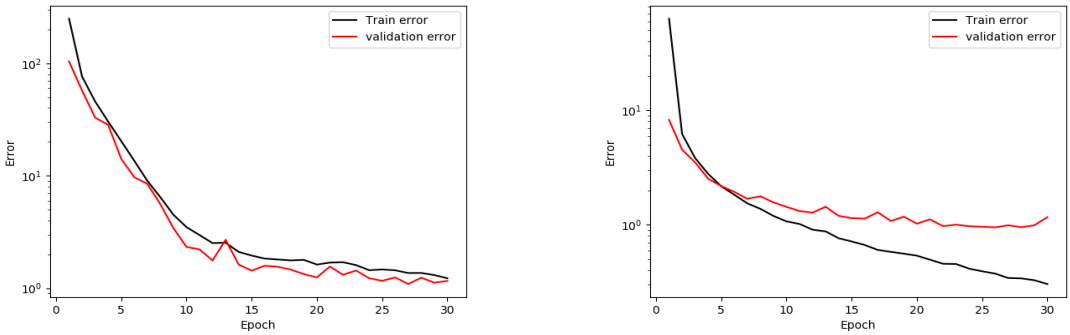


(a) Training Curve (xy)



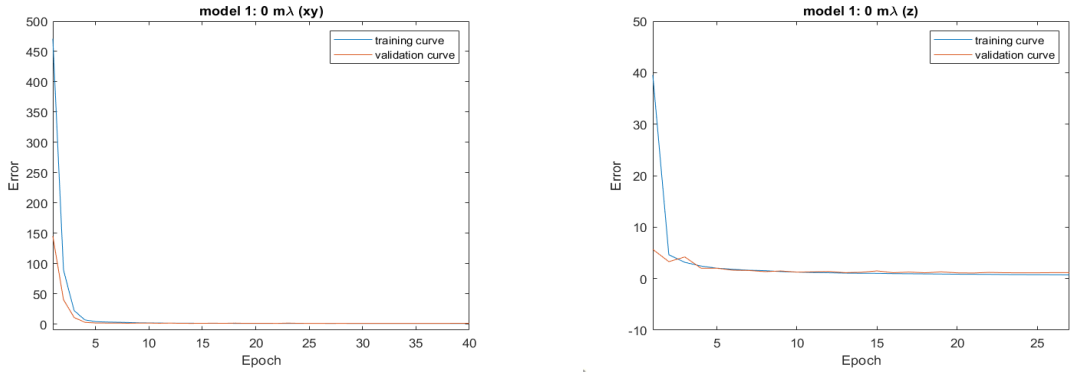
(b) Training Curve (z)

Figure 8.6: Training Curve : Experiment 2 - 3D localization when signal photon and background are constant



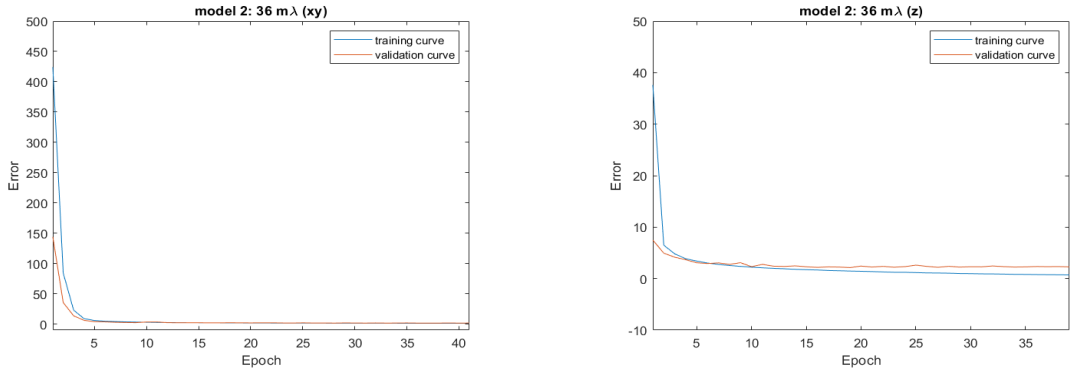
(a) Training Curve (xy) (b) Training Curve (z)

Figure 8.7: Training Curve: Experiment 3 - Training smNet with varying signal photon count and background



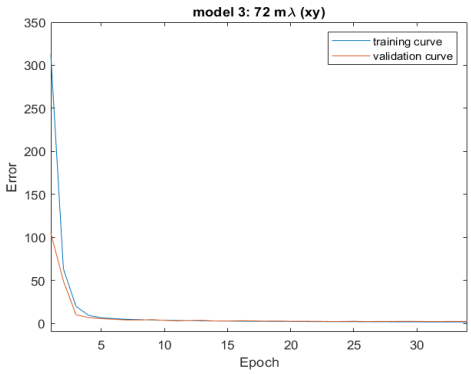
(a) Training Curve (xy) (b) Training Curve (z)

Figure 8.8: Training Curve of model 1 - splitting the parameter space

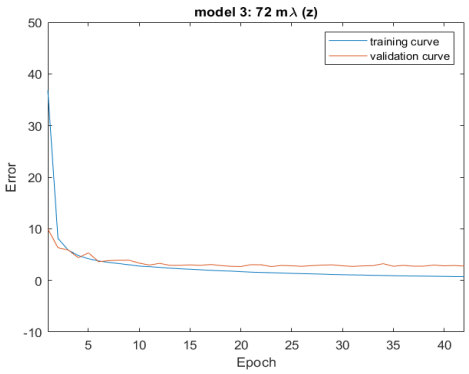


(a) Training Curve (xy) (b) Training Curve (z)

Figure 8.9: Training Curve of model 2 - splitting the parameter space

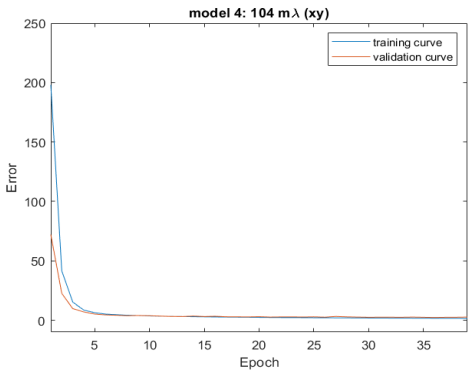


(a) Training Curve (xy)

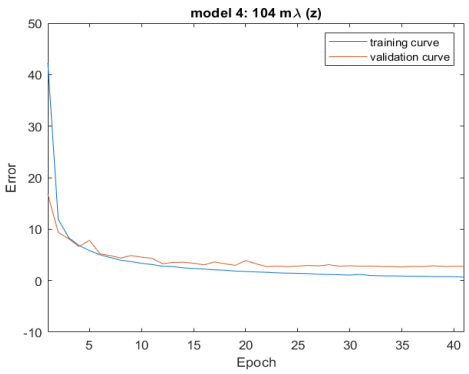


(b) Training Curve (z)

Figure 8.10: Training Curve of model 3 - splitting the parameter space



(a) Training Curve (xy)



(b) Training Curve (z)

Figure 8.11: Training Curve of model 4 - splitting the parameter space

Bibliography

[1] W Patrick Ambrose, Peter M Goodwin, Jörg Enderlein, David J Semin, John C Martin, and Richard A Keller. Fluorescence photon antibunching from single molecules on a surface. *Chemical Physics Letters*, 269(3):365 – 370, 1997. ISSN 0009-2614. doi: https://doi.org/10.1016/S0009-2614(97)00266-2. URL http://www.sciencedirect.com/science/article/pii/S0009261497002662.

[2] Sean B. Andersson. Localization of a fluorescent source without numerical fitting. *Opt. Express*, 16(23):18714–18724, Nov 2008. doi: 10.1364/OE.16.018714. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-16-23-18714.

[3] Paolo Annibale, Stefano Vanni, Marco Scarselli, Ursula Rothlisberger, and Aleksandra Radenovic. Quantitative photo activated localization microscopy: Unraveling the effects of photoblinking. *PLOS ONE*, 6(7):1–8, 07 2011. doi: 10.1371/journal.pone.0022678. URL https://doi.org/10.1371/journal.pone.0022678.

[4] Toshimitsu Aritake, Hideitsu Hino, Shigeyuki Namiki, Daisuke Asanuma, Kenzo Hirose, and Noboru Murata. Fast and robust multiplane single molecule localization microscopy using deep neural network, 2020.

[5] Hazen Babcock, Yaron M. Sigal, and Xiaowei Zhuang. A high-density 3d localization algorithm for stochastic optical reconstruction microscopy. *Optical Nanoscopy*, 1(1):6, 2012. ISSN 2192-2853. doi: 10.1186/2192-2853-1-6. URL https://doi.org/10.1186/2192-2853-1-6.

[6] David Baddeley, Mark B. Cannell, and Christian Soeller. Visualization of localization microscopy data. *Microscopy and Microanalysis*, 16(1):64–72, 2010. doi: 10.1017/S143192760999122X.

[7] Mark R. Baker and Rajendra B. Patil. Universal approximation theorem for interval neural networks. *Reliable Computing*, 4(3):235–239, Aug 1998. ISSN 1573-1340. doi: 10.1023/A:1009951412412. URL https://doi.org/10.1023/A:1009951412412.

[8] Martin J. Booth. *Adaptive Optics in Microscopy*, chapter 14, pages 295–322. John Wiley Sons, Ltd, 2011. ISBN 9783527635245. doi: 10.1002/9783527635245.ch14. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527635245.ch14.

[9] Benjamin P. Bowen, Allan Scruggs, Jörg Enderlein, Markus Sauer, and Neal Woodbury. Implementation of neural networks for the identification of single molecules. *The Journal of Physical Chemistry A*, 108(21):4799–4804, 2004. doi: 10.1021/jp036456v. URL https://doi.org/10.1021/jp036456v.

[10] Nicholas Boyd, Eric Jonas, Hazen Babcock, and Benjamin Recht. Deeploco: Fast 3d localization microscopy using neural networks. *bioRxiv*, 2018. doi: https://www.biorxiv.org/content/10.1101/267096v1. URL https://www.biorxiv.org/content/10.1101/267096v1.

[11] C R Burch and J P P Stock. Phase-contrast microscopy. *Journal of Scientific Instruments*, 19 (5):71–75, may 1942. doi: 10.1088/0950-7671/19/5/302. URL https://doi.org/10.1088%2F0950-7671%2F19%2F5%2F302.

[12] Roxana Busca. Master thesis: Wavefront reconstruction from intensity based images for high resolution microscopy. Delft University of Technology, November 2016.

[13] Jerry Chao, E. Sally Ward, and Raimund J. Ober. Fisher information theory for parameter estimation in single molecule microscopy: tutorial. *J. Opt. Soc. Am. A*, 33(7):B36–B57, Jul 2016. doi: 10.1364/JOSAA.33.000B36. URL http://josaa.osa.org/abstract.cfm?URI=josaa-33-7-B36.

[14] Jules R. Dim and Tamio Takamura. Alternative approach for satellite cloud classification: Edge gradient application. *Advances in Meteorology*, 2013:584816, Dec 2013. ISSN 1687-9309. doi: 10.1155/2013/584816. URL https://doi.org/10.1155/2013/584816.

[15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 184–199, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10593-2.

[16] Alexander Egner, Claudia Geisler, Claas von Middendorff, Hannes Bock, Dirk Wenzel, Rebecca Medda, Martin Andresen, Andre C. Stiel, Stefan Jakobs, Christian Eggeling, Andreas Schönle, and Stefan W. Hell. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical journal*, 93(9):3285–3290, Nov 2007. ISSN 0006-3495. doi: 10.1529/biophysj.107.112201. URL https://pubmed.ncbi.nlm.nih.gov/17660318. 17660318[pmid].

[17] Jörg Enderlein, Richard A. Keller, and Christoph Zander. *Single Molecule Detection in Liquids and on Surfaces under Ambient Conditions: Introduction and Historical Overview*, chapter 1, pages 1–19. John Wiley and Sons, Ltd, 2003. ISBN 9783527600809. doi: 10.1002/3527600809.ch1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/3527600809.ch1.

[18] Philippe Fischer and Greg P. Kochanski. Optimal Addition of Images for Detection and Photometry. *Astronomical Journal*, 107:802, Feb 1994. doi: 10.1086/116898.

[19] Claudia Geisler, Thomas Hotz, Andreas Schönle, Stefan W. Hell, Axel Munk, and Alexander Egner. Drift estimation for single marker switching based imaging schemes. *Opt. Express*, 20(7):7274–7289, Mar 2012. doi: 10.1364/OE.20.007274. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-20-7-7274.

[20] Song Han, Jeff Pool, John Tran, and William J. Dally. Learning both weights and connections for efficient neural networks. *CoRR*, abs/1506.02626, 2015. URL http://arxiv.org/abs/1506.02626.

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[22] Ricardo Henriques, Mickael Lelek, Eugenio F. Fornasiero, Flavia Valtorta, Christophe Zimmer, and Musa M. Mhlanga. Quickpalm: 3d real-time photoactivation nanoscopy image processing in imagej. *Nature Methods*, 7(5):339–340, 2010. ISSN 1548-7105. doi: 10.1038/nmeth0510-339. URL https://doi.org/10.1038/nmeth0510-339.

[23] Seamus J. Holden, Stephan Uphoff, and Achillefs N. Kapanidis. Daostorm: an algorithm for high-density super-resolution microscopy. *Nature Methods*, 8(4):279–280, 2011. ISSN 1548-7105. doi: 10.1038/nmeth0411-279. URL https://doi.org/10.1038/nmeth0411-279.

[24] Eelco Hoogendoorn, Kevin C. Crosby, Daniela Leyton-Puig, Ronald M. P. Breedijk, Kees Jalink, Theodorus W. J. Gadella, and Marten Postma. The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. *Scientific Reports*, 4(1):3854, Jan 2014. ISSN 2045-2322. doi: 10.1038/srep03854. URL https://doi.org/10.1038/srep03854.

[25] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (New York, N.Y.)*, 319(5864):810–813, Feb 2008. ISSN 1095-9203. doi: 10.1126/science.1153529. URL https://pubmed.ncbi.nlm.nih.gov/18174397. 18174397[pmid].

[26] Fang Huang, Samantha L. Schwartz, Jason M. Byars, and Keith A. Lidke. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomed. Opt. Express*, 2(5):1377–1393, May 2011. doi: 10.1364/BOE.2.001377. URL http://www.osapublishing.org/boe/abstract.cfm?URI=boe-2-5-1377.

[27] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015. URL http://arxiv.org/abs/1502.03167.

[28] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

[29] Teresa Klein, Sven Proppert, and Markus Sauer. Eight years of single-molecule localization microscopy. *Histochemistry and Cell Biology*, 141(6):561–575, Jun 2014. ISSN 1432-119X. doi: 10.1007/s00418-014-1184-3. URL https://doi.org/10.1007/s00418-014-1184-3.

[30] Wouter M. Kouw. An introduction to domain adaptation and transfer learning. *CoRR*, abs/1812.11806, 2018. URL http://arxiv.org/abs/1812.11806.

[31] Pavel Křížek, Ivan Raška, and Guy M. Hagen. Minimizing detection errors in single molecule localization microscopy. *Opt. Express*, 19(4):3226–3235, Feb 2011. doi: 10.1364/OE.19.003226. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-19-4-3226.

[32] Joshua D. Larkin and Peter R. Cook. Maximum precision closed-form solution for localizing diffraction-limited spots in noisy images. *Opt. Express*, 20(16):18478–18493, Jul 2012. doi: 10.1364/OE.20.018478. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-20-16-18478.

[33] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. ISSN 1476-4687. doi: 10.1038/nature14539. URL https://doi.org/10.1038/nature14539.

[34] Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. Pruning filters for efficient convnets. *CoRR*, abs/1608.08710, 2016. URL http://arxiv.org/abs/1608.08710.

[35] Yiming Li, Yuji Ishitsuka, Per Niklas Hedde, and G. Ulrich Nienhaus. Fast and efficient molecule detection in localization-based super-resolution microscopy by parallel adaptive histogram equalization. *ACS Nano*, 7(6):5207–5214, Jun 2013. ISSN 1936-0851. doi: 10.1021/nn4009388. URL https://doi.org/10.1021/nn4009388.

[36] Yiming Li, Markus Mund, Philipp Hoess, Joran Deschamps, Ulf Matti, Bianca Nijmeijer, Vilma Jimenez Sabinina, Jan Ellenberg, Ingmar Schoen, and Jonas Ries. Real-time 3d single-molecule localization using experimental point spread functions. *Nature Methods*, 15(5):367–369, 2018. ISSN 1548-7105. doi: 10.1038/nmeth.4661. URL https://doi.org/10.1038/nmeth.4661.

[37] Sheng Liu and Keith A. Lidke. A multiemitter localization comparison of 3d superresolution imaging modalities. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 15(4):696–704, Mar 2014. ISSN 1439-7641. doi: 10.1002/cphc.201300758. URL https://pubmed.ncbi.nlm.nih.gov/24281982. 24281982[pmid].

[38] Zhuang Liu, Jianguo Li, Zhiqiang Shen, Gao Huang, Shoumeng Yan, and Changshui Zhang. Learning efficient convolutional networks through network slimming. *CoRR*, abs/1708.06519, 2017. URL http://arxiv.org/abs/1708.06519.

[39] Hongqiang Ma, Jianquan Xu, Jingyi Jin, Ying Gao, Li Lan, and Yang Liu. Fast and precise 3d fluorophore localization based on gradient fitting. *Scientific Reports*, 5, 2015. doi: https://doi.org/10.1038/srep14335.

[40] Koen J. A. Martens, Arjen N. Bader, Sander Baas, Bernd Rieger, and Johannes Hohlbein. Phasor based single-molecule localization microscopy in 3d (psmlm-3d): An algorithm for mhz localization rates using standard cpus. *The Journal of Chemical Physics*, 148(12):123311, 2018. doi: 10.1063/1.5005899. URL https://doi.org/10.1063/1.5005899.

[41] Gleb Vdovin Michel Verhaegen and Oleg Soloviev. Control for high resolution imaging: Lecture notes for the course sc4045. Delft University of Technology, April 2016.

[42] Junhong Min, Cédric Vonesch, Hagai Kirshner, Lina Carlini, Nicolas Olivier, Seamus Holden, Suliana Manley, Jong Chul Ye, and Michael Unser. Falcon: fast and unbiased reconstruction of high-density super-resolution microscopy data. *Scientific Reports*, 4(1):4577, 2014. ISSN 2045-2322. doi: 10.1038/srep04577. URL https://doi.org/10.1038/srep04577.

[43] Michael J. Mlodzianoski, John M. Schreiner, Steven P. Callahan, Katarina Smolková, Andrea Dlasková, Jitka Šantorová, Petr Ježek, and Joerg Bewersdorf. Sample drift correction in 3d fluorescence photoactivation localization microscopy. *Opt. Express*, 19(16):15009–15019, Aug 2011. doi: 10.1364/OE.19.015009. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-19-16-15009.

[44] Leonhard Mockl, Anish R. Roy, Petar N. Petrov, and W. E. Moerner. Accurate and rapid background estimation in single-molecule localization microscopy using the deep neural network bgnet. *Proceedings of the National Academy of Sciences*, 117(1):60–67, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1916219117. URL https://www.pnas.org/content/117/1/60.

[45] Elias Nehme, Lucien E. Weiss, Tomer Michaeli, and Yoav Shechtman. Deep-storm: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458–464, Apr 2018. doi: 10.1364/OPTICA.5.000458. URL http://www.osapublishing.org/optica/abstract.cfm?URI=optica-5-4-458.

[46] Elias Nehme, Daniel Freedman, Racheli Gordon, Boris Ferdman, Lucien E. Weiss, Onit Alalouf, Reut Orange, Tomer Michaeli, and Yoav Shechtman. Deepstorm3d: dense three dimensional localization microscopy and point spread function design by deep learning. *arXiv: Image and Video Processing*, 2019. URL https://arxiv.org/abs/1906.09957.

[47] Robert Nieuwenhuizen. Doctoral thesis: Quantitative image analysis for single molecule localization microscopy. Delft University of Technology, 2016. URL https://doi.org/10.4233/uuid:f7c64c64-08e9-4e18-9834-72264f45509e.

[48] Robert P. J. Nieuwenhuizen, Keith A. Lidke, Mark Bates, Daniela Leyton Puig, David Grünwald, Sjoerd Stallinga, and Bernd Rieger. Measuring image resolution in optical nanoscopy. *Nature Methods*, 10(6):557–562, Jun 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2448. URL https://doi.org/10.1038/nmeth.2448.

[49] Robert J. Noll. Zernike polynomials and atmospheric turbulence*. *J. Opt. Soc. Am.*, 66(3):207–211, Mar 1976. doi: 10.1364/JOSA.66.000207. URL http://www.osapublishing.org/abstract.cfm?URI=josa-66-3-207.

[50] Raimund J. Ober, Sripad Ram, and E. Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, Feb 2004. ISSN 0006-3495. doi: 10.1016/S0006-3495(04)74193-4. URL https://pubmed.ncbi.nlm.nih.gov/14747353.14747353[pmid].

[51] Jean-Christophe Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989 – 1996, 2002. ISSN 0031-3203. doi: https://doi.org/10.1016/S0031-3203(01)00127-3. URL http://www.sciencedirect.com/science/article/pii/S0031320301001273.

[52] Wei Ouyang, Andrey Aristov, Mickaël Lelek, Xian Hao, and Christophe Zimmer. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology*, 36(5):460–468, 2018. ISSN 1546-1696. doi: 10.1038/nbt.4106. URL https://doi.org/10.1038/nbt.4106.

[53] Raghuveer Parthasarathy. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nature Methods*, 9(7):724–726, 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2071. URL https://doi.org/10.1038/nmeth.2071.

[54] Alexandros Pertsinidis, Yunxiang Zhang, and Steven Chu. Subnanometre single-molecule localization, registration and distance measurements. *Nature*, 466(7306):647–651, Jul 2010. ISSN 1476-4687. doi: 10.1038/nature09163. URL https://doi.org/10.1038/nature09163.

[55] Tingwei Quan, Hongyu Zhu, Xiaomao Liu, Yongfeng Liu, Jiuping Ding, Shaoqun Zeng, and Zhen-Li Huang. High-density localization of active molecules using structured sparse model and bayesian information criterion. *Opt. Express*, 19(18):16963–16974, Aug 2011. doi: 10.1364/OE.19.016963. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-19-18-16963.

[56] C. Radhakrishna Rao. Linear statistical inference and its applications. *Wiley Series in Probability and Statistics.*, December 2001.

[57] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

[58] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. URL http://arxiv.org/abs/1409.1556.

[59] Carlas Smith, Max Huisman, Marijn Siemons, David Grünwald, and Sjoerd Stallinga. Simultaneous measurement of emission color and 3d position of single molecules. *Opt. Express*, 24(5):4996–5013, Mar ts. doi: 10.1364/OE.24.004996. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-24-5-4996.

[60] Carlas S. Smith, Nikolai Joseph, Bernd Rieger, and Keith A. Lidke. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nature Methods*, 7(5):373–375, 2010. ISSN 1548-7105. doi: 10.1038/nmeth.1449. URL https://doi.org/10.1038/nmeth.1449.

[61] Artur Speiser, Srinivas C. Turaga, and Jakob H. Macke. Teaching deep neural networks to localize sources in super-resolution microscopy by combining simulation-based learning and unsupervised learning. *CoRR*, abs/1907.00770, 2019. URL http://arxiv.org/abs/1907.00770.

[62] Sjoerd Stallinga and Jeroen Kalkman. Imaging system: Lecture notes for the course ap3121. Delft University of Technology, April 2018.

[63] Sjoerd Stallinga and Bernd Rieger. Accuracy of the gaussian point spread function model in 2d localization microscopy. *Opt. Express*, 18(24):24461–24476, Nov ts. doi: 10.1364/OE.18.024461. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-18-24-24461.

[64] Peter B. Stetson. DAOPHOT - a computer program for crowded-field stellar photometry. *Publications of the Astronomical Society of the Pacific*, 99:191, mar 1987. doi: 10.1086/131977. URL https://doi.org/10.1086%2F131977.

[65] T. Takeshima, T. Takahashi, J. Yamhasita, Y. Okada, and S. Watanabe. A multi-emitter fitting algorithm for potential live cell super-resolution imaging over a wide range of molecular densities. *Journal of Microscopy*, 271(3):266–281, 2018. doi: 10.1111/jmi.12714. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jmi.12714.

[66] Anisha Thayil and Martin J. Booth. Self calibration of sensorless adaptive optical microscopes. 2011.

[67] Rasmus Ø. Thorsen, Christiaan N. Hulleman, Mathias Hammer, David Grünwald, Sjoerd Stallinga, and Bernd Rieger. Impact of optical aberrations on axial position determination by photometry. *Nature Methods*, 15(12):989–990, Dec 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0227-4. URL https://doi.org/10.1038/s41592-018-0227-4.

[68] Yina Wang, Tingwei Quan, Shaoqun Zeng, and Zhen-Li Huang. Palmer: a method capable of parallel localization of multiple emitters for high-density localization microscopy. *Opt. Express*, 20(14): 16039–16049, Jul 2012. doi: 10.1364/OE.20.016039. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-20-14-16039.

[69] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W. Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12):1090–1097, 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0216-7. URL https://doi.org/10.1038/s41592-018-0216-7.

[70] Weidi Xie, J. Alison Noble, and Andrew Zisserman. Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):283–292, 2018. doi: 10.1080/21681163.2016.1149104. URL https://doi.org/10.1080/21681163.2016.1149104.

[71] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, Aug 1995. ISSN 1939-3539. doi: 10.1109/34.400568.

[72] Andrew G. York, Alireza Ghitani, Alipasha Vaziri, Michael W. Davidson, and Hari Shroff. Confined activation and subdiffractive localization enables whole-cell palm with genetically expressed probes. *Nature Methods*, 8(4):327–333, 2011. ISSN 1548-7105. doi: 10.1038/nmeth.1571. URL https://doi.org/10.1038/nmeth.1571.

[73] C. Zander, K.H. Drexhage, K.-T. Han, J. Wolfrum, and M. Sauer. Single-molecule counting and identification in a microcapillary. *Chemical Physics Letters*, 286(5):457 – 465, 1998. ISSN 0009-2614. doi: https://doi.org/10.1016/S0009-2614(98)00096-7. URL http://www.sciencedirect.com/science/article/pii/S0009261498000967.

[74] P. Zelger, K. Kaser, B. Rossboth, L. Velas, G. J. Schütz, and A. Jesacher. Three-dimensional localization microscopy using deep learning. *Opt. Express*, 26(25):33166–33179, Dec 2018. doi: 10.1364/OE.26.033166. URL http://www.opticsexpress.org/abstract.cfm?URI=oe-26-25-33166.

[75] Frits Zernike. Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica*, (1):689–704, May 1934.

[76] Peiyi Zhang, Sheng Liu, Abhishek Chaurasia, Donghan Ma, Michael J. Mlodzianoski, Eugenio Culurciello, and Fang Huang. Analyzing complex single-molecule emission patterns with deep learning. *Nature Methods*, 15(11):913–916, 2018. ISSN 1548-7105. doi: 10.1038/s41592-018-0153-5. URL https://doi.org/10.1038/s41592-018-0153-5.