A Machine Learning Method for Object Localization

Mridul Gupta, Mary L. Comer, Edward J. Delp
Video and Image Processing Lab (VIPER)
School of Electrical and Computer Engineering
Purdue University
Jonathan Chan, Mitchell Krouss, Paul Martens,
Michael Jacobs, Corbin F. Spells, Moses W. Chan
Advanced Technology Center
Lockheed Martin Corporation

ABSTRACT

In many applications, the spatial resolution of the imaging system is not enough to localize a small object or multiple closely-spaced objects (CSOs). Detection and localization of objects are more difficult if they are small in size (occupy a few pixels) and have a low signal-to-noise ratio (SNR). We model the noise as an additive noise using a zero-mean Gaussian distribution with identical variance for all the pixels. We assume we know the approximate 2D spatial neighborhood containing the objects in an observed image. We also consider the case with two spatially registered spectral bands where the bands have different ensquared energy, which is the fraction of energy on a detector when an object lies at its center. For our experiments, we assume that the imaging system's point spread function (PSF) is a 2D symmetric Gaussian function with its center at the object's location and width dependent on the ensquared energy of the spectral band. Sub-pixel location estimator for small objects (SPLEO) is a lightweight convolutional neural network-based method that estimates an object's sub-pixel location and amplitude in an observed image. In this paper, we propose an improvement to SPLEO called SPLEOv2 for localizing CSOs. We show that by using the parameter estimates (spatial locations and amplitudes) from SPLEOv2 as initial estimates for 3 existing approaches for object localization, we can achieve a better mean, standard deviation, and 95^{th} percentile localization error. We also study the variation in localization performance of SPLEOv2 when information from two spectral bands is available. We compare the variance in the parameter estimates with the Cramér-Rao lower bound (CRLB) on the variance of an unbiased estimator.

1. INTRODUCTION AND RELATED WORK

Detection and sub-pixel localization of objects have applications across several domains, such as resolving binary stars in astronomy [4, 11, 34], monitoring protein interactions in microscopy [18, 17, 32, 1, 5, 16, 29, 30, 10], and object tracking [27, 26, 12, 15, 13, 14] for reconnaissance and surveillance. A single pixel in these applications usually represents a significantly larger area than the objects of interest. Most methods [27, 26, 12, 15, 18, 2, 3, 8, 24] that have been developed for this problem assume that the point spread function (PSF) of the imaging system is known, and the object is modeled as a point source with a spatial shape dependent on the PSF. Localizing an object becomes difficult when the signal-to-noise ratio (SNR) [12] is low, as the noise interferes with the spatial shape of the object.

Several compressed sensing methods [18, 34] have been developed, which model this problem as a discrete deconvolution problem (Equation 1).

$$\vec{S} = K\vec{x_0} + \vec{\varepsilon}, \tag{1}$$

where $\vec{S}$ is a row-concatenated (flattened) vector of the observed image, K is a shape matrix dependent on the PSF, and $\vec{\varepsilon}$ is the noise. $\vec{x_0}$ is a vector with non-zero elements at locations where an object is present. $\vec{x_0}$ is determined by minimizing the difference between $\vec{S}$ and $K\vec{x_0}$. Sparsity is promoted in $\vec{x_0}$ by adding a penalty term such as l1

norm [34] and l0 norm [18]. These methods assume that the PSF of the imaging system is known, which determines the shape matrix K in Equation 1. Since compressed sensing methods use discrete deconvolution to localize objects, their resolution capability has a discrete limit.

Reagan et al. [27] proposed a method for localizing a single object in an image with a single spectral band using maximum likelihood estimation. It minimizes the cost function shown in Equation 2 using gradient descent [7] to determine the sub-pixel location and amplitude of an object present in the observed image.

$$\min_{x_1,y_1,A_1} ||\vec{S} - A_1\vec{g}(x_1,y_1)||^2, \tag{2}$$

where $\vec{S}$ is a row-concatenated (flattened) vector of the observed image, A_1 is the object amplitude, and $\vec{g}(x_1,y_1)$ is the flattened object shape matrix which is dependent on the point spread function (PSF) of the imaging system and the location of the object (x_1,y_1). Equation 2 is differentiated with respect to A_1 and equated with 0 to obtain A_1 in terms of $\vec{S}$ and $\vec{g}(x_1,y_1)$. The cost function shown in Equation 3 is obtained by substituting A_1 in Equation 2 and is minimized using gradient descent [7] to estimate (x_1,y_1).

$$\min_{x_1,y_1}\left[-\frac{(\vec{S}^T\vec{g}(x_1,y_1))^2}{\vec{g}^T(x_1,y_1)\vec{g}(x_1,y_1)}\right]. \tag{3}$$

Another method was proposed by Reagan et al. [26] for localizing two closely-spaced objects (CSOs) in an observed image with a single spectral band using maximum likelihood estimation. For two CSOs, Equation 2 and Equation 3 are updated as shown in Equation 4 and Equation 5, respectively.

$$\min_{\vec{x},\vec{y},\vec{A}} ||\vec{S} - G(\vec{x},\vec{y})\vec{A}||^2, \tag{4}$$

$$\min_{\vec{x},\vec{y}} -(\vec{S}^T G(\vec{x},\vec{y})[G^T(\vec{x},\vec{y})G(\vec{x},\vec{y})]^{-1}G^T(\vec{x},\vec{y})\vec{S}), \tag{5}$$

where $\vec{x} = [x_1,x_2]$, $\vec{y} = [y_1,y_2]$, and $G(\vec{x},\vec{y}) = [\vec{g}(x_1,y_1)\ \ \vec{g}(x_2,y_2)]$. (x_1,y_1), (x_2,y_2), and $\vec{A} = [A_1,A_2]$ are the locations and amplitudes of the two CSOs.

Gupta et al. [12] proposed a method called Dual Band Object Localization (DBL) that uses information from two spectral bands to localize a single object or two CSOs. For different spectral bands, the PSF of the imaging system is assumed to be a 2D symmetric Gaussian function with different widths. The maximum likelihood-based cost function shown in Equation 6 is minimized using the Nelder Mead non-linear optimization method [23, 21] to estimate the sub-pixel locations and amplitudes of the objects present in the observed image.

$$\min_{\vec{x},\vec{y},\vec{A}} ||\vec{S}_{B_1} - \vec{S}_{e,B_1}||^2 + ||\vec{S}_{B_2} - \vec{S}_{e,B_2}||^2, \tag{6}$$

where $\vec{S}_{B_k}$ is the row-concatenated (flattened) observed image in the spectral band B_k. For a single object $\vec{x} = [x_1]$,$\vec{y} = [y_1]$, and $\vec{A} = [A_{1,B_1},A_{1,B_2}]$. (x_1,y_1) is the spatial location of the object, A_{1,B_k} is the amplitude of the object in the spectral band B_k, and $\vec{S}_{e,B_k} = A_{1,B_k}\ \vec{g}(x_1,y_1,B_k)$ is the flattened expected observed image (without noise) in the spectral band B_k given the sub-pixel location and amplitudes of the object. Since the width of the PSF varies for spectral bands, the spatial shape of the object $(\vec{g}(x_1,y_1,B_k))$ is dependent on both object location and spectral band. Similarly, for two CSOs, $\vec{x} = [x_1,x_2]$,$\vec{y} = [y_1,y_2]$, and $\vec{A} = [A_{1,B_1},A_{1,B_2},A_{2,B_1},A_{2,B_2}]$. (x_1,y_1) and (x_2,y_2) are the spatial locations of the two objects, A_{t,B_k} is the amplitude of the t^{th} object in the spectral band B_k, and $\vec{S}_{e,B_k} = A_{1,B_k}\ \vec{g}(x_1,y_1,B_k) + A_{2,B_k}\ \vec{g}(x_2,y_2,B_k)$.

Deep learning approaches [6, 20, 15] have also been proposed for object detection and localization, but most of them are not expected to perform well for localizing or detecting point sources [9]. These methods require a large amount of training data that is often unavailable. Gupta et al. proposed a deep learning approach called Sub-Pixel Location Estimator for Small Objects (SPLEO) [15], which requires limited training data and can localize point sources efficiently at low signal-to-noise ratios (SNRs) [12, 15]. SPLEO localizes a single object present in an image using a single spectral band. It processes a single spectral band image with a deep neural network that outputs an object's

spatial sub-pixel location and amplitude. The network is trained by minimizing a cost function derived from the maximum-likelihood estimation of the observed image (Equation 2).

In this paper, we propose an improvement to SPLEO called SPLEOv2 for localizing a single object or two closely-spaced objects (CSOs) using observed images of size 7×7 in one or two spatially registered spectral bands. SPLEOv2 can be easily modified to localize more than two CSOs as well. We assume that the image formation process is known, the point spread function (PSF) of the imaging system is a 2D symmetric Gaussian function, and that we know the objects' approximate spatial locations. The network is trained by minimizing the observed image's negative log-likelihood and estimates the objects' sub-pixel locations and amplitudes.

2. MODEL OF IMAGE FORMATION PROCESS

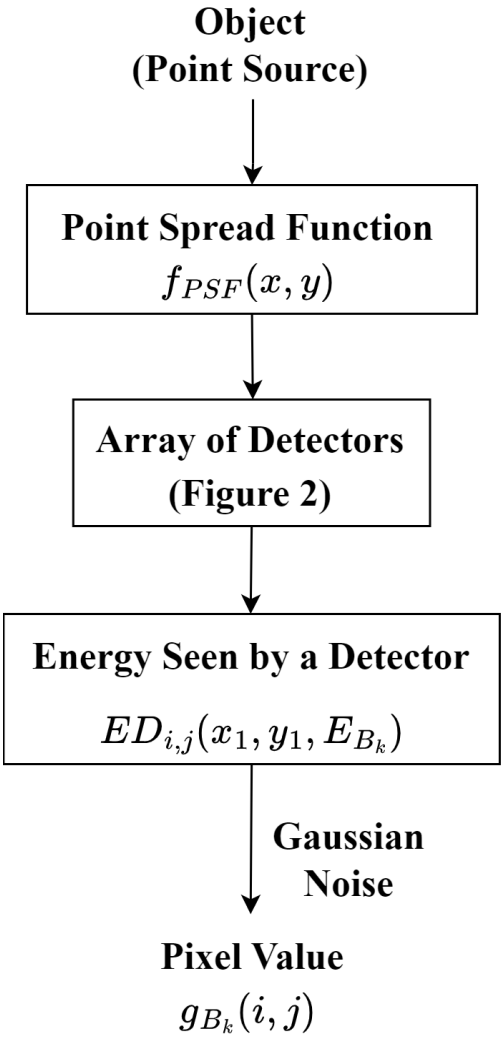

Fig. 1: Model of the image formation process. (Source: Gupta et al. [12])

Here we describe our model of image capture and formation process as shown in Fig. 1. The objects are modeled as point sources which are processed by the imaging system to form the observed images. As shown in Fig. 1, the energy due to the point source is spread over an array of detectors because of the point spread function (PSF) of the imaging system. After the PSF, the energy on a detector is integrated and noise is added to determine the respective pixel value in the observed image. The imaging system's PSF is assumed to be a 2D symmetric Gaussian function with width

σ_{PSF} (Equation 7).

$$f_{PSF}(x,y) = \frac{1}{2\pi\sigma_{PSF}^2} e^{-\frac{1}{2\sigma_{PSF}^2}\left(x^2+y^2\right)}. \tag{7}$$

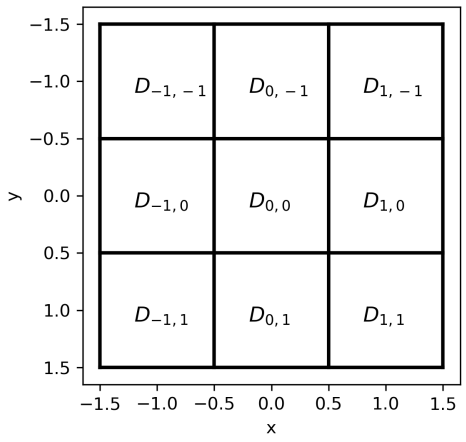

Fig. 2: Detector array where each $D_{i,j}$ is a detector, and the center of each detector is a pixel in the observed image. (Source: Gupta et al. [12])

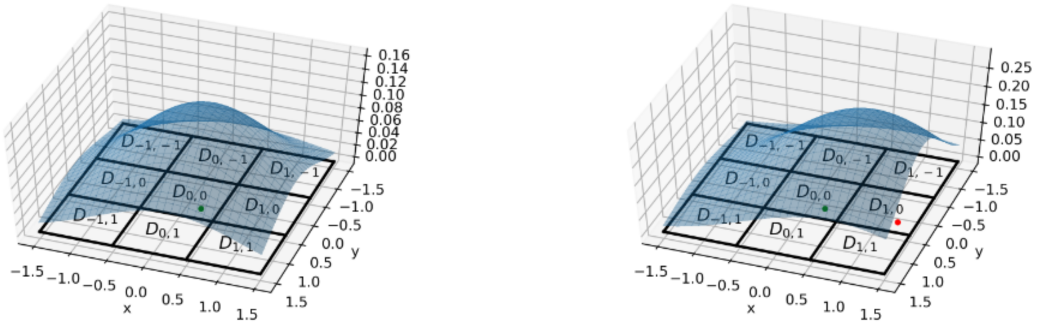

Fig. 3: Energy of a single object (left) and two CSOs (right) is spread over multiple detectors due to the point spread function (PSF). Object locations are marked in red and green. (Source: Gupta et al. [12])

An array of detectors is shown in Fig. 2. The center of a detector $D_{i,j}$ represents the pixel at location (i, j) in the observed image. A pixel value is determined by integrating the energy over the corresponding detector in the detector array. The width of the PSF is usually large enough to spread the energy due to an object to nearby detectors, as shown in Fig. 3. The energy seen by a detector can be contributed by multiple objects, as shown in Fig. 3.

For an object with amplitude 1 and spatial location (x_1, y_1), Equation 8 determines how the energy is spread in its spatial neighborhood.

$$f_{1obj}(x,y) = \frac{1}{2\pi\sigma_{PSF}^2} e^{-\frac{1}{2\sigma_{PSF}^2}\left((x-x_1)^2+(y-y_1)^2\right)}. \tag{8}$$

If there are two objects with unit amplitude at spatial locations (x_1, y_1) and (x_2, y_2), a detector may see the energy from

both objects, as shown in Fig. 3 (Equation 9).

$$f_{2obj}(x,y) = \frac{1}{2\pi\sigma_{PSF}^2} e^{-\frac{1}{2\sigma_{PSF}^2}\left((x-x_1)^2+(y-y_1)^2\right)} + \frac{1}{2\pi\sigma_{PSF}^2} e^{-\frac{1}{2\sigma_{PSF}^2}\left((x-x_2)^2+(y-y_2)^2\right)}. \tag{9}$$

The energy on a detector $D_{0,0}$ when an object is present at location (x_1, y_1) is given by Equation 10.

$$ED_{0,0} = \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} f_{1obj}(x,y)dx\,dy, \tag{10}$$

where -0.5 to 0.5 is the span of the detector $D_{0,0}$ in x and y directions.

For 2 CSOs located at (x_1, y_1) and (x_2, y_2) the energy on a detector $D_{0,0}$ is added from both objects as shown in Equation 11.

$$ED_{0,0} = \int_{-0.5}^{0.5}\int_{-0.5}^{0.5} f_{2obj}(x,y)dx\,dy. \tag{11}$$

Similarly, we can determine the energy on all detectors due to the objects present in an image.

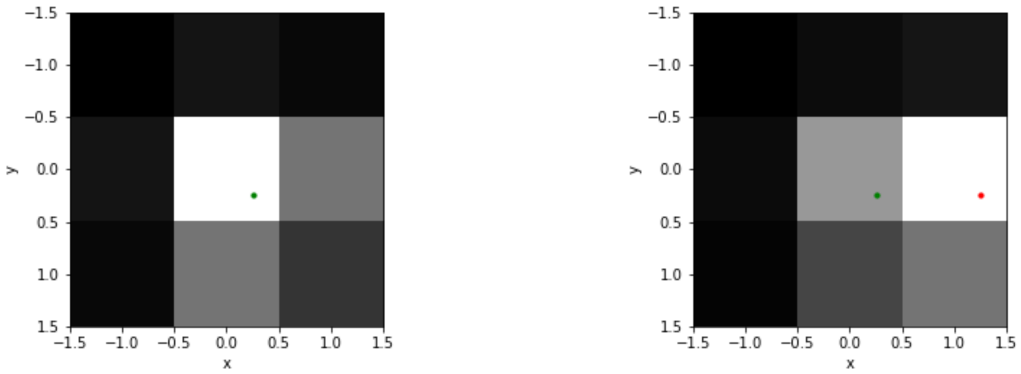

Fig. 4: Simulated observed images with a single object (left) and two CSOs (right). The center of the observed images is at $(0,0)$. Object locations are marked in red and green. (Source: Gupta et al. [12])

We consider two spatially registered spectral bands in our experiments and assume that the PSF of both spectral bands is a symmetric 2D Gaussian function with different widths. The width of the PSF for band B_k is $\sigma_{PSF_{B_k}}$. Ensquared energy is the ratio of energy collected on a detector when an object lies at its center. We denote ensquared energy for a band B_k with E_{B_k}, and it is related to the width of the PSF by Equation 12.

$$\sigma_{PSF_{B_k}} = \frac{1}{2\sqrt{2}\,erf^{-1}(\sqrt{E_{B_k}})}, \tag{12}$$

where

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt. \tag{13}$$

Substituting Equation 12 in Equation 10, energy on a detector can be written as a function of ensquared energy (Equation 14).

$$ED_{0,0} = \frac{1}{2}\left(erf\Big[(0.5-x_1)\alpha\Big] - erf\Big[(-0.5-x_1)\alpha)\Big]\right) \times \frac{1}{2}\left(erf\Big[(0.5-y_1)\alpha\Big] - erf\Big[(-0.5-y_1)\alpha\Big]\right), \tag{14}$$

where

$$\alpha = 2 \times erf^{-1}(\sqrt{E_{B_k}}), \tag{15}$$
$$erf(-x) = -erf(x).$$

The energy on a detector $D_{i,j}$ in the spectral band B_k due to an object at spatial location (x_1, y_1) can be computed using Equation 16.

$$ED_{i,j}(x_1, y_1, E_{B_k}) = \frac{1}{2}\left(erf\left[(u+0.5)\alpha\right] - erf\left[(u-0.5)\alpha\right]\right) \times \frac{1}{2}\left(erf\left[(v+0.5)\alpha\right] - erf\left[(v-0.5)\alpha\right]\right), \quad (16)$$

where

$$u = i - x_1,$$
$$v = j - y_1.$$

We model the noise using a zero mean Gaussian distribution with identical variance for all pixels ($\varepsilon \sim \mathcal{N}(0, \sigma^2_{noise})$). For an object present at spatial location (x_1, y_1), the pixel value ($g_{B_k}(i, j)$) corresponding to the detector $D_{i,j}$ in the spectral band B_k is given by Equation 17.

$$g_{B_k}(i, j) = ED_{i,j}(x_1, y_1, E_{B_k}) + \varepsilon_{i,j}, \quad (17)$$

where $\varepsilon_{i,j}$ is the Gaussian noise at location (i, j). We derived the expression for pixel value in Equation 17 for objects with unit amplitude. For an object at location (x_1, y_1) with amplitude A_{1,B_k} in the spectral band B_k, we can compute the pixel value at location (i, j) by scaling the energy seen by the detector with the amplitude of the object as shown in Equation 18.

$$g_{B_k}(i, j) = A_{1,B_k} ED_{i,j}(x_1, y_1, E_{B_k}) + \varepsilon_{i,j}. \quad (18)$$

Similarly, for two objects at locations (x_1, y_1) and (x_2, y_2) with amplitudes A_{1,B_k} and A_{2,B_k} in spectral band B_k, respectively, pixel value at location (i, j) can computed using Equation 19.

$$g_{B_k}(i, j) = A_{1,B_k} D_{i,j}(x_1, y_1, E_{B_k}) + A_{2,B_k} D_{i,j}(x_2, y_2, E_{B_k}) + \varepsilon_{i,j}. \quad (19)$$

Fig. 4. shows observed images with one object and two CSOs simulated using the process described above. For our experiments, we simulate images of size 7×7.

3. PROPOSED APPROACH

Fig. 5 shows a block diagram of our proposed approach Sub-Pixel Location Estimator for Small Objects v2 (SPLEOv2). Similar to SPLEO [15], SPLEOv2 employs a lightweight convolutional network. The model is lightweight as it does not require a dedicated graphics processing unit (GPU) for training and has a low computation time [31, 33]. It has approximately 0.9 million parameters which are significantly less than 30 million parameters in U-Net [28]. SPLEOv2 requires 0.0001 seconds to process one observed image of size 7×7 with Python on an Intel Core i9 $-$ 9900X 3.50 GHz CPU with 128 GB RAM. We assume that we know the approximate location of the objects in the observed image and that the observed image is formed using the process described in Section 2. We examine a window of size 7×7 around the approximate object location to estimate the objects' spatial sub-pixel locations and amplitudes.

The input to SPLEOv2 is a $C \times 7 \times 7$ size tensor referred to as the observed window denoted by S, and C represents the number of spectral bands. The input tensor is processed by an average pooling layer with window size 2×2 and two 2D convolutional layers having 32 filters each of sizes 3×3 and 5×5, respectively. In Fig. 5, Conv2D(f, m) represents a 2D convolution layer with f filters of size $m \times m$. The outputs of the two 2D convolutional layers are then processed with average pooling layers with window size 2×2. The outputs from the three average pooling layers are then flattened and concatenated to obtain a single 1D vector which is then processed with 3 fully connected layers having $512, 20$, and p neurons, respectively. A fully connected layer with n neurons is represented by FC(n) in Fig. 5. Rectified linear unit (ReLU) [22] is used as the activation function for all layers except the last layer, which uses hyperbolic tangent function (Tanh) [22]. The output of the network is a vector ($\vec{P} = [P_1, P_2, ...P_p]$) with p elements in the range $(-1, 1)$ which is the range of Tanh function. p is the number of parameters to be estimated, which depends on the number of objects present in the image and the number of spectral bands. For a single object in a single spectral band, the parameters to be estimated are the spatial location and amplitude of the object ($p = 3$). For a single object in two spectral bands, $p = 4$ because we need to estimate the amplitude of the object in both spectral bands. Similarly, $p = 6$ for two objects and one spectral band while $p = 8$ for two objects and two spectral bands. Our experiments assume that the objects are located within the central 3×3 window of the observed image $x_1, y_1, x_2, y_2 \in (-1.5, 1.5)$.

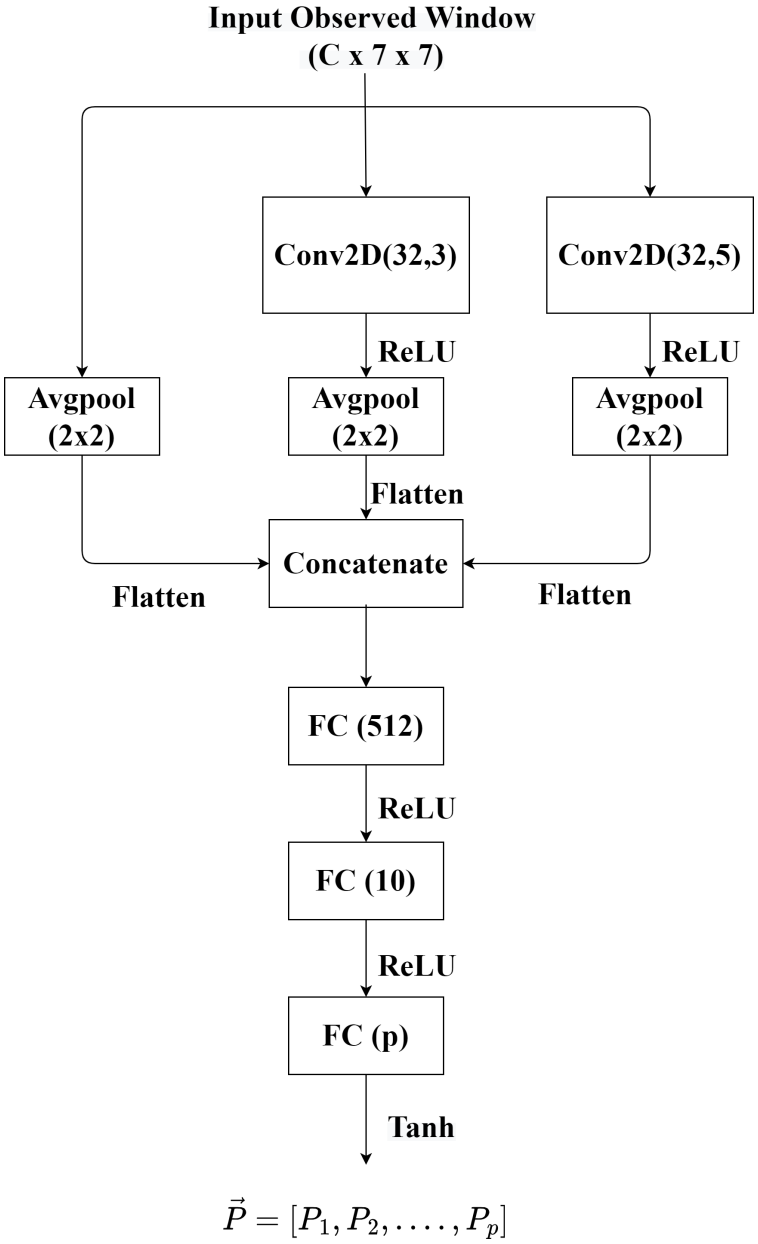

Fig. 5: Block diagram of the proposed method Sub-Pixel Location Estimator for Small Objects v2 (SPLEOv2). Input is a $C \times 7 \times 7$ size tensor where C is the number of spectral bands. Conv2D(f,m) represents a 2D convolution layer with f filters of size $m \times m$, FC(n) represents a fully connected layer with n neurons, AvgPool is an average pooling layer with window size 2×2, ReLU is rectified linear unit, and Tanh is a hyperbolic tangent function.

The amplitude of an object can be greater than 1 so we divide the pixel values in the input observed images with a normalization factor N_f, which is greater than the maximum possible amplitude for any object. For a single object in one spectral band, the output vector $\vec{P}$ is related to the location and amplitude estimates as shown in Equation 20.

$$x_{e_1} = 3P_1, \;\; y_{e_1} = 3P_2, \;\; A_{e_1,B_1} = N_f|P_3|, \tag{20}$$

where x_{e_i}, y_{e_i}, and A_{e_i,B_k} are the location and amplitude estimates of the i^{th} object in spectral band B_k. This means the network can estimate object locations in the range $(-3,3)$ and amplitude in the range $(0,N_f)$. Equation 21 shows the

relationship between output vector and parameter estimates for a single object in two spectral bands.

$$x_{e_1} = 3P_1,\ \ y_{e_1} = 3P_2,\ \ A_{e_1,B_1} = N_f|P_3|,\ \ A_{e_1,B_2} = N_f|P_3|(1.5|P_4|+0.5), \tag{21}$$

where $A_{e_1,B_2} = N_f|P_3|(1.5|P_4|+0.5)$ ensures that the object's amplitude ratio in the two bands is more than 0.5 and less than 2. Equation 22 shows the relationship between output vector and parameter estimates for two CSOs in a single spectral band.

$$x_{e_1} = 3P_1,\ \ y_{e_1} = 3P_2,\ \ x_{e_2} = 3P_3,\ \ y_{e_2} = 3P_4,\ \ A_{e_1,B_1} = N_f|P_5|,\ \ A_{e_2,B_1} = N_f|P_5|(1.5|P_6|+0.5), \tag{22}$$

where the ratio of the amplitude of the two objects is more than 0.5 and less than 2. Equation 23 shows the relationship between output vector and parameter estimates for two CSOs in two spectral bands.

$$\begin{aligned} &x_{e_1} = 3P_1,\ \ y_{e_1} = 3P_2,\ \ x_{e_2} = 3P_3,\ \ y_{e_2} = 3P_4,\ \ A_{e_1,B_1} = N_f|P_5|, \\ &A_{e_2,B_1} = N_f|P_5|(1.5|P_6|+0.5),\ \ A_{e_1,B_2} = N_f|P_5|(1.5|P_7|+0.5),\ \ A_{e_2,B_2} = N_f|P_5|(1.5|P_8|+0.5), \end{aligned} \tag{23}$$

where the ratio of any two amplitudes is more than 0.5 and less than 2. Note that this restriction on the ratio of the amplitudes is necessary for all comparison methods (including SPLEOv2) to perform well.

Now, we will derive the cost function that will be minimized to train SPLEOv2. Given the estimates for locations and amplitudes of the objects, we can construct the expected observed window of size 7×7 in the absence of noise. We call this reconstructed observed window the expected observed window (EOW) denoted by S_e. For a single object present at location (x_1, y_1), the EOW in a spectral band B_k can be estimated as shown in Equation 24.

$$S_{e,B_k} = A_{e_1,B_k}\ S'(x_{e_1}, y_{e_1}, B_k), \tag{24}$$

where $S'(x_{e_1}, y_{e_1}, B_k)$ is computed in a manner similar to $ED_{i,j}(x_1, y_1, E_{B_k})$ (Equation 25):

$$S'_{i,j}(x_{e_1}, y_{e_1}, B_k) = \frac{1}{2}\left(erf\left[(u+0.5)\alpha\right] - erf\left[(u-0.5)\alpha\right]\right) \times \frac{1}{2}\left(erf\left[(v+0.5)\alpha\right] - erf\left[(v-0.5)\alpha\right]\right), \tag{25}$$

where

$$\begin{aligned} \alpha &= 2 \times erf^{-1}(\sqrt{E_{B_k}}), \\ u &= i - x_{e_1}, \\ v &= j - y_{e_1}, \end{aligned} \tag{26}$$

and $i, j \in \{-3, -2, ..., 3\}$ are the pixel locations in the observed window.

Similarly, for two closely-spaced objects, Equation 27 can be used to estimate the expected observed window.

$$S_{e,B_k} = A_{e_1,B_k}\ S'(x_{e_1}, y_{e_1}, B_k) + A_{e_2,B_k}\ S'(x_{e_2}, y_{e_2}, B_k). \tag{27}$$

The conditional probability distribution of the pixel values of the observed window S given the estimates for the objects' spatial locations and amplitudes is given by Equation 28.

$$f(S/\vec{x}, \vec{y}, \vec{A}) = \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{(\vec{S}-\vec{S_e})^T(\vec{S}-\vec{S_e})}{2\sigma_g^2}}, \tag{28}$$

where $\vec{Q}$ represents a row-concatenated (flattened) matrix Q. The negative log-likelihood of the pixel values after removing the terms independent of location and amplitude parameters is given by Equation 29. Note that σ_g is identical for all pixels and is independent of the location and amplitude parameters.

$$-\log f(S/\vec{x}, \vec{y}, \vec{A}) = (\vec{S}-\vec{S_e})^T(\vec{S}-\vec{S_e}). \tag{29}$$

We minimize Equation 29 using the Adam optimizer [19], batch size of 100, and a learning rate of 0.001.

4. EXPERIMENTAL RESULTS

To evaluate our proposed approach Sub-Pixel Location Estimator for Small Objects v2 (SPLEOv2), we conducted experiments with 7×7 size observed windows simulated using the process described in Section 2. Recall that we model the noise using a zero-mean Gaussian distribution with variance σ^2_{noise}. σ^2_{noise} is related to the signal-to-noise ratio [12] as shown in Equation 30.

$$SNR_{dB} = 10log_{10}\frac{P_{signal}}{\sigma_{noise}}, \tag{30}$$

where P_{signal} is the maximum energy seen by a detector (Equation 31).

$$P_{signal} = \max_{i,j} ED_{i,j}(0,0,E_{B_k}). \tag{31}$$

Equation 32 is used to compute σ_{noise} for $SNR_{dB} \in \{2,4,6,8,10,12,14\}$.

$$\sigma_{noise} = \frac{\max_{i,j} ED_{i,j}(0,0,E_{B_k})}{10^{\frac{SNR_{dB}}{10}}}. \tag{32}$$

Observed windows with a single spectral band are simulated for $E_{B_1} \in \{0.3,0.4,0.5,0.6,0.7\}$ and observed windows with two spectral bands are simulated for $(E_{B_1},E_{B_2})] \in \{(0.3,0.3),(0.4,0.4),(0.5,0.5),(0.6,0.6),(0.7,0.7)\}$.

We compare the performance of SPLEOv2 with existing approaches using the metrics: mean localization error, the standard deviation of localization error, 95^{th} percentile error of localization error, and normalized mean amplitude error. Mean localization error is the mean of the Euclidean distance between the true location (x_1,y_1) of an object and its estimated location (x_{e_1},y_{e_1}). Localization error can be computed as shown in Equation 33.

$$l_{e_1} = \sqrt{(x_1 - x_{e_1})^2 + (y_1 - y_{e_1})^2}. \tag{33}$$

Normalized mean amplitude error is the mean of the absolute difference between the true amplitude (A_{1,B_k}) and the estimated amplitude (A_{e_1,B_k}) normalized by the true amplitude. Normalized amplitude error can be computed as shown in Equation 34.

$$e_A = \frac{|A_{1,B_k} - A_{e_1,B_k}|}{A_{1,B_k}}. \tag{34}$$

We also compare the variance in our parameter estimates (spatial locations and amplitudes of the objects) with the Cramér-Rao Lower Bound (CRLB) [25] on the variance of an unbiased estimator. If r independent observations of a random variable x with probability density function $f(x;b)$ are used to estimate the parameter b, then Cramér-Rao inequality is given by Equation 35.

$$var(\hat{b}) \geq \frac{1}{r\,I(b)}, \tag{35}$$

where $\hat{b}$ is an estimate of parameter b, $var(\hat{b})$ is the variance of $\hat{b}$, $\frac{1}{r\,I(b)}$ is the CRLB [25] on the variance of $\hat{b}$, and $I(b)$ is defined as shown in Equation 36.

$$I(b) = -E\left[\frac{\partial^2}{\partial b^2} ln(f(x;b))\right]. \tag{36}$$

Note the CRLB is the lower bound for an unbiased estimator, so a biased estimator can perform better than this lower bound. By biased, we mean that for a parameter b, $E[\hat{b}] \neq b$. A derivation of Cramér-Rao Lower Bound (CRLB) can be found in [12].

4.1 Observed Windows Containing a Single Object

To compare the performance of our approach with existing approaches, we simulated 13,500 observed windows of size 7×7 containing a single object using the process described in Section 2. The object is placed randomly according to a uniform distribution over the central 3×3 window ($x_1,y_1 \in ((-1.5,1.5))$). For observed windows with a single spectral band, the images are simulated for each pair of $SNR_{dB} \in \{2,4,6,8,10,12,14\}$ and $E_{B_1} \in$

$\{0.3, 0.4, 0.5, 0.6, 0.7\}$. For observed windows with two spectral bands, the images are simulated for each pair of $SNR_{dB} \in \{2, 4, 6, 8, 10, 12, 14\}$ and $(E_{B_1}, E_{B_2}) \in \{(0.3, 0.3), (0.4, 0.4), (0.5, 0.5), (0.6, 0.6), (0.7, 0.7)\}$. The amplitude of the objects varies uniformly in the range $(1, 10)$ and is equal for both spectral bands. Recall from Section 3 that the observed windows are divided by a normalization factor (N_f) before they are processed with SPLEOv2. For our experiments with a single object, we set $N_f = 20$.

We trained SPLEOv2 for various ensquared energy and SNRs and determined that the model trained on high SNR images generalizes well to low SNR images. We present the localization results obtained by training the model on $10,000$ observed images of size 7×7 having $18dB$ SNR and an ensquared energy of 0.5 for both spectral bands. The object is placed randomly according to a uniform distribution over the central 3×3 window $(x_1, y_1 \in ((-1.5, 1.5)))$ and the amplitude varies in the range $(1, 10)$ for both bands.

We compare SPLEOv2 with 4 existing approaches [15, 27, 12]. SPLEO [15] is a deep learning method that uses a single spectral band to localize a single object in an observed window. It assumes that the number of objects in the image and the imaging system's point spread function (PSF) are known. It is trained on high SNR images and generalizes well to low SNR images. We also compare our method's performance with the observed image's centroid (location estimate). We do not obtain an amplitude estimate with centroiding, so we only compare its results using localization error. For images with two spectral bands, the centroid is computed for the mean of the two spectral bands. Reagan et al. [27] minimizes a maximum-likelihood-based cost function using gradient descent [7] to localize a single object in an observed image with a single spectral band. It also assumes that the number of objects and PSF of the imaging system are known. DBL [12] minimizes a maximum-likelihood-based cost function using Nelder-Mead non-linear optimization [23, 21] to localize a single object or two closely-spaced objects in an observed image with two spectral bands. It assumes that the PSF of the imaging system is known but does not need to know the number of objects. Reagan et al. [27] and DBL [12] use non-linear optimization, so they require an initial estimate for the locations and amplitudes of the objects. We consider two cases for the initial estimates. In the first case, the centroid is taken as the initial estimate for the spatial location and the ensquared energy of the band as the initial estimate for the amplitudes. In the second case, outputs from SPLEO and SPLEOv2 are processed as shown in Equation 20 and Equation 21 and considered as the initial estimates for location and amplitudes. We represent results for Reagan et al. [27] and DBL [12] with Reagan et al. $[27]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods.

Fig. 6 shows the mean, standard deviation, and 95^{th} percentile localization error variation with signal-to-noise ratio (SNR) [12] and ensquared energy. Recall that we simulate $13,500$ observed windows for each pair of SNR and Ensquared energy. The mean, standard deviation and 95^{th} percentile localization error are computed over these $13,500$ windows. The object lies within the central 3×3 window $(x_1, y_1 \in (-1.5, 1.5))$. Variation with SNR_{dB} is shown at an ensquared energy of 0.5 and variation with ensquared energy is shown at an SNR of $6dB$. Centroiding does not provide a good location estimate when the object lies outside the central pixel. Reagan et al. [27] and DBL [12] perform better with SPLEO [15] and SPLEOv2 as initial estimates, respectively. While SPLEOv2 has a slightly higher mean localization error than DBL, it has a comparable standard deviation and 95^{th} percentile localization error. DBL has the lowest mean localization error at all SNRs when it uses SPLEOv2 as its initial estimate. DBL also has a significantly lower standard deviation and 95^{th} percentile localization error when it uses SPLEOv2 as its initial estimate. This is because SPLEOv2 is a good estimate of an object's location which is then refined by DBL using non-linear optimization. At higher SNRs, all methods except centroiding have similar performance as localization becomes easier. Methods using two spectral bands perform better than those using a single spectral band because more information is available for localizing objects. Mean, standard deviation, and 95^{th} percentile localization error do not vary significantly for most methods with a change in ensquared energy. Due to centroiding becoming a worse estimate for location, Reagan et al. [27] 's and DBL's performance with centroid as initial estimate worsens with increased ensquared energy. DBL with SPLEOv2 as the initial estimate has the best performance, as observed from the three comparison metrics for most SNR and all ensquared energy values. Fig. 7 shows the normalized mean amplitude error variation with signal-to-noise ratio (SNR) [12] and the ensquared energy. The object lies within the central 3×3 window $(x_1, y_1 \in (-1.5, 1.5))$. Results are similar for the amplitudes in the two spectral bands. Fig. 7 shows results for amplitude estimates in only the first spectral band. Similar to localization error, variations with SNR and ensquared energy are shown for ensquared energy of 0.5 and SNR of $6dB$, respectively. Amplitude errors are similar for all methods at all SNRs. It gets higher with an increase in ensquared energy because most of the energy is concentrated within a single pixel leading to a higher pixel value, making the methods estimate a higher amplitude than the true amplitude.

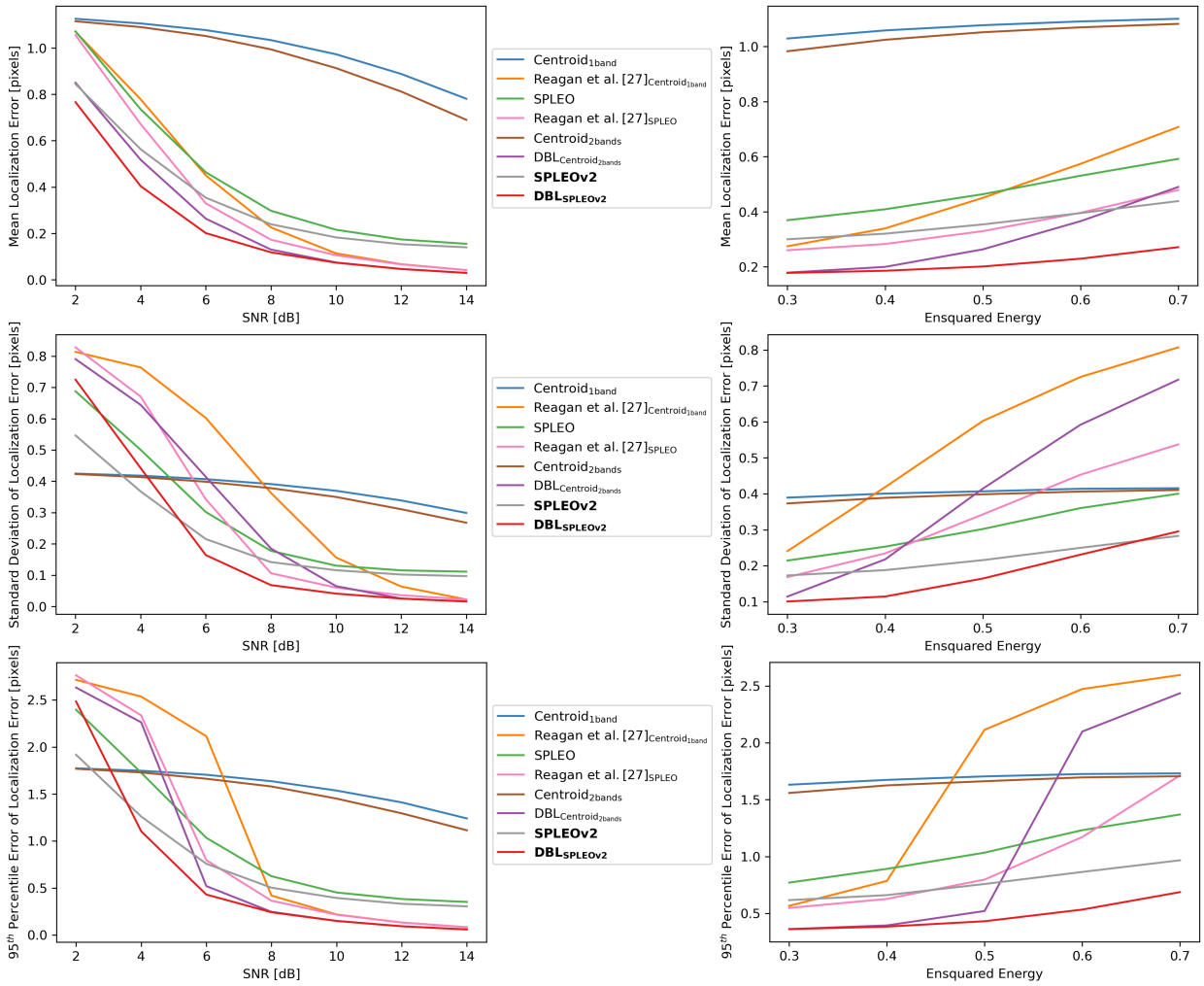

Fig. 6: Variation of mean, standard deviation, and 95^{th} percentile localization error with SNR (left) and ensquared energy (right). We represent results for Reagan et al. [27] and DBL [12] with Reagan et al. $[27]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods. There is a single object in the observed window.

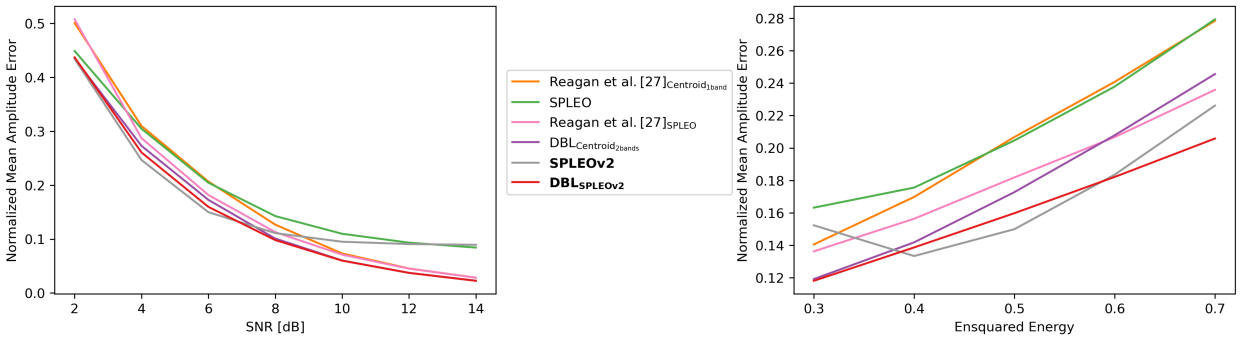

Fig. 7: Variation of normalized mean amplitude error with SNR (left) and ensquared energy (right). We represent results for Reagan et al. [27] and DBL [12] with Reagan et al. $[27]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods. There is a single object in the observed window.

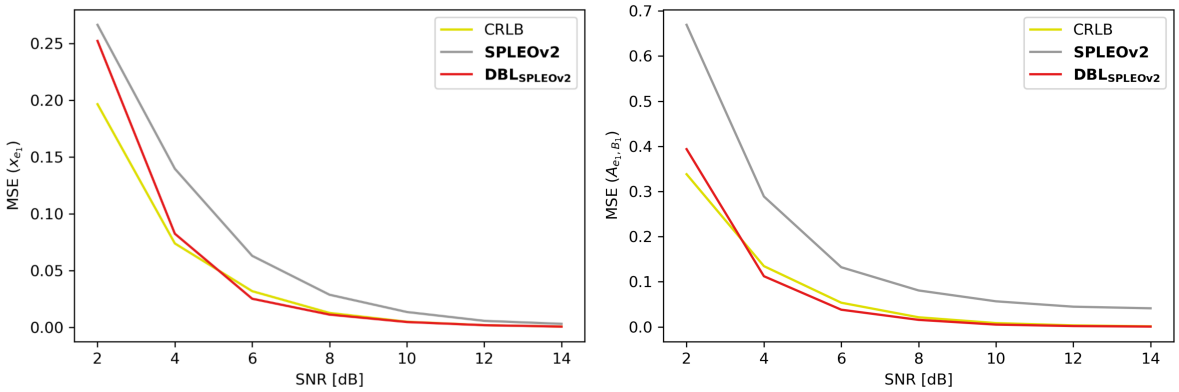

Fig. 8: Comparison of mean squared error in parameter estimates obtained from SPLEOv2 with Cramér-Rao Lower Bound. There is a single object at the center of the observed window. Observed window has 2 spectral bands.

Fig. 8 shows a comparison of mean squared error (MSE) in the parameter estimates (x_{e_1} and A_{e_1,B_1}) with the Cramér-Rao Lower Bound (CRLB) [25]. MSE for x_{e_1} and y_{e_1} are similar and MSE for A_{e_1,B_1} and A_{e_1,B_2} are similar. We generate 500 windows of size 7×7 with ensquared energy 0.5 for both spectral bands. The object is located at the center of the observed windows ($x_1 = 0, y_1 = 0$). SPLEOv2 is unable to achieve the CRLB, but DBL with SPLEOv2 as an initial estimate performs similarly to the CRLB. As mentioned before, DBL refined the estimates obtained from SPLEOv2 using non-linear optimization, making its performance better than SPLEOv2. Performance is similar to CRLB at high SNRs as the localization becomes easier, and CRLB gets close to 0.

4.2 Observed Windows Containing Two Closely-Spaced Objects (CSOs)

To compare the performance of our approach with existing approaches, we simulated $10,000$ observed windows of size 7×7 containing two closely-spaced objects (CSOs) using the process described in Section 2. The first object lies within the central pixel ($x_1, y_1 \in (-0.5, 0.5)$), and the second object is $d \in \{0.1, 0.2, ..., 1\}$ pixels away from the first object. The relative orientation of the two objects is distributed uniformly in the range $(0^\circ - 360^\circ)$. We generate $1,000$ observed windows for each value of d. For observed windows with a single spectral band, the images are simulated for each pair of $SNR_{dB} \in \{2, 4, 6, 8, 10, 12, 14\}$ and $E_{B_1} \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$. For observed windows with two spectral bands, the images are simulated for each pair of $SNR_{dB} \in \{2, 4, 6, 8, 10, 12, 14\}$ and $(E_{B_1}, E_{B_2})] \in \{(0.3, 0.3), (0.4, 0.4), (0.5, 0.5), (0.6, 0.6), (0.7, 0.7)\}$. The amplitude of the objects varies uniformly in the range $(1, 10)$ and is equal for both objects in both spectral bands. Recall from Section 3 that the pixel values in the observed windows are divided by a normalization factor (N_f) before they are processed with SPLEOv2. For our experiments with CSOs, we set $N_f = 40$.

Similar to a single object case, we trained SPLEOv2 for various ensquared energy and SNRs and determined that the model trained on high SNR images generalizes well to low SNR images. We present the localization results obtained by training the model on $60,000$ observed images of size 7×7 having $18dB$ SNR and an ensquared energy of 0.5 for both spectral bands. The first object lies within the central pixel ($x_1, y_1 \in (-0.5, 0.5)$), and the second object is $d \in \{0.1, 0.2, ..., 1\}$ pixels away from the first object. The relative orientation of the two objects is distributed uniformly in the range $(0^\circ - 360^\circ)$. Training data comprises of $6,000$ observed windows for each value of $d \in \{0.1, 0.2, ..., 1\}$.

We compare SPLEOv2 with 2 existing approaches [26, 12]. Reagan et al. [26] minimizes a maximum-likelihood-based cost function using gradient descent [7] to localize two CSOs in an observed image with a single spectral band. It assumes that the number of objects and PSF of the imaging system are known. DBL [12] minimizes a maximum-likelihood-based cost function using Nelder-Mead non-linear optimization [23, 21] to localize a single object or two closely-spaced objects in an observed image with two spectral bands. It assumes that the PSF of the imaging system is known but does not need to know the number of objects. Reagan et al. [26] and DBL [12] use non-linear optimization, so they require an initial estimate for the locations and amplitudes of the objects. We consider two cases for the initial estimates. In the first case, the two brightest pixels are taken as the initial estimate for the spatial location and the ensquared energy of the band as the initial estimate for the amplitudes. In the second case, outputs from SPLEOv2

are processed as shown in Equation 22 and Equation 23 and considered as the initial estimates for locations and amplitudes. We represent results for Reagan et al. [26] and DBL [12] with Reagan et al. $[26]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods.

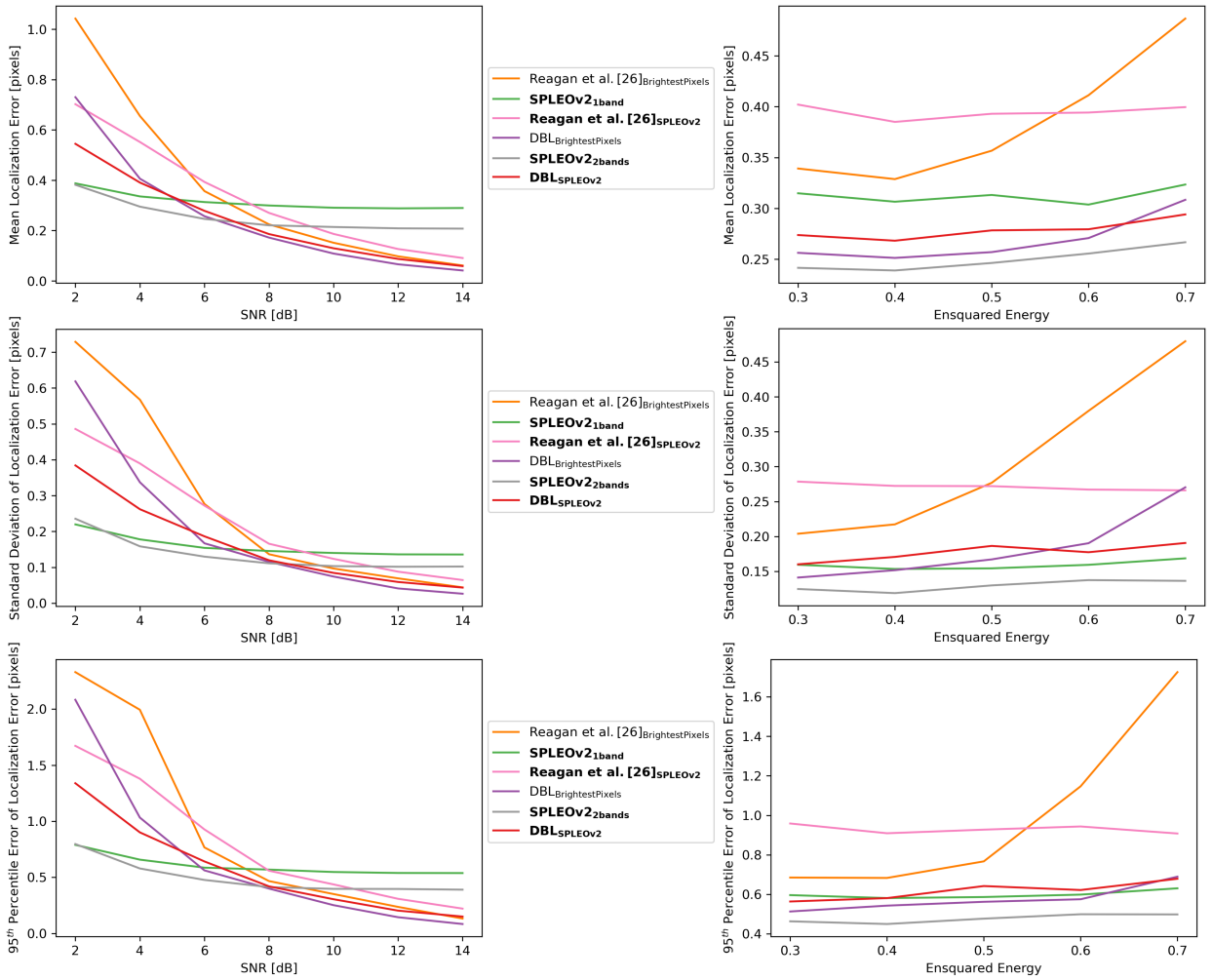

Fig. 9: Variation of mean, standard deviation, and 95^{th} percentile localization error with SNR (left) and ensquared energy (right). We represent results for Reagan et al. [26] and DBL [12] with Reagan et al. $[26]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods. There are two objects in the observed window. Results are shown for the object within the central pixel.

Fig. 9 shows the mean, standard deviation, and 95^{th} percentile localization error variation with signal-to-noise ratio (SNR) [12] and ensquared energy. Recall that we simulate $1,000$ windows for each value of d, the distance between the two pixels. Mean, standard deviation and 95^{th} percentile localization error are estimated for these $1,000$ windows. The two objects are 0.7 pixels apart, with one object lying within the central pixel. The results are shown for the object within the central pixel and are similar for the other object. Variation with SNR is shown at an ensquared energy of 0.5 and variation with ensquared energy is shown at an SNR of $6dB$. Reagan et al. [26] and DBL [12] perform better with SPLEOv2 as an initial estimate at low SNRs. SPLEOv2 has the lowest mean, standard deviation, and 95^{th} percentile localization error for SNR $< 6dB$. Its performance stays similar with an increase in SNR. DBL has the lowest mean, standard deviation, and 95^{th} percentile localization error at all SNRs $\geq 6dB$ when it uses SPLEOv2 as its initial estimate. This is because SPLEOv2 is a good estimate of an object's location which is then refined by DBL using non-linear optimization. At higher SNRs, all methods have similar performance as localization becomes easier. Methods using two spectral bands perform better than those using a single spectral band because more information is available for localizing objects. Mean, standard deviation and 95^{th} percentile localization error do not vary significantly for

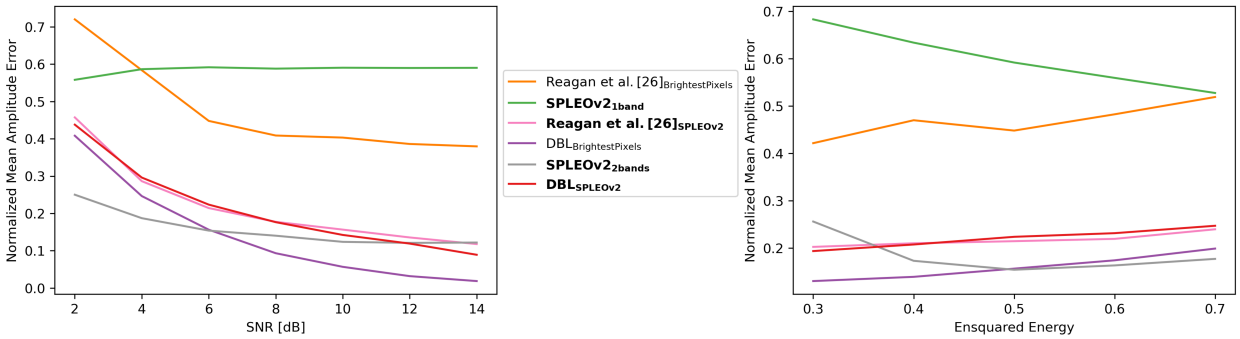

Fig. 10: Variation of normalized mean amplitude error with SNR (left) and ensquared energy (right). We represent results for Reagan et al. [26] and DBL [12] with Reagan et al. $[26]_{\text{init}}$ and DBL_{init} where init is the initial estimate for these methods. There are two objects in the observed window. Results are shown for the object within the central pixel.

most methods with a change in ensquared energy. SPLEOv2 and DBL with SPLEOv2 as the initial estimate perform consistently well, as observed from the three comparison metrics for most SNR and all ensquared energy values. Fig. 10 shows the normalized mean amplitude error variation with signal-to-noise ratio (SNR) [12] and the ensquared energy. The two objects are 0.7 pixels apart, with one object lying within the central pixel. The results are shown for the object within the central pixel and are similar for the other object. Results are similar for the amplitudes in the two spectral bands, so we present results for amplitude in one spectral band. Similar to localization error, variation with SNR is shown at an ensquared energy of 0.5 and variation with ensquared energy is shown at an SNR of $6dB$. Ensquared energy is a bad estimate for amplitude, making Reagan et al. [27] amplitude bad. The initial estimate for amplitude does not affect the performance when we have two spectral bands. DBL and SPLEOv2 have similar performance for amplitudes when two spectral bands are used. Normalized mean amplitude error does not vary much with a change in ensquared energy.

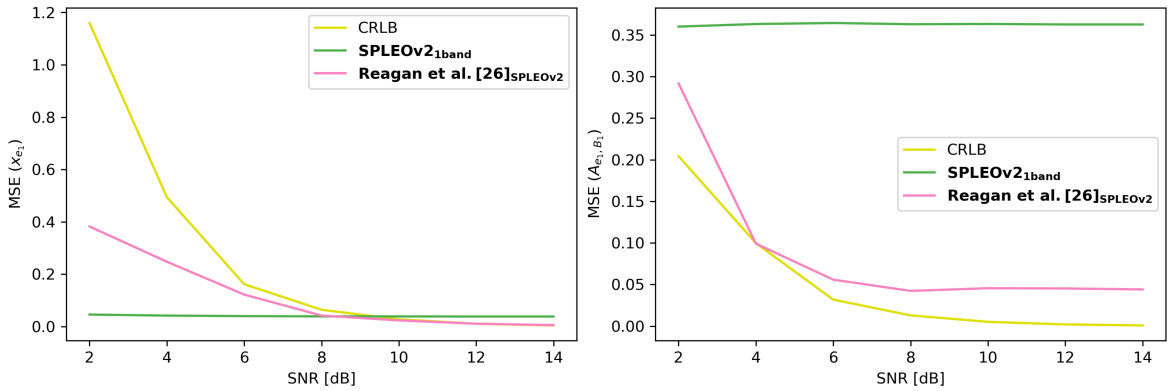

Fig. 11: Comparison of mean squared error in parameter estimates obtained from SPLEOv2 with Cramér-Rao Lower Bound. One object is at the center of the observed window. The second object is 0.7 pixels away along the x-axis. Results are shown for the first object. The observed window has 1 spectral band.

Fig. 11 shows a comparison of mean squared error (MSE) in the parameter estimates (x_{e_1} and A_{e_1,B_1}) obtained from SPLEOv2 with the Cramér-Rao Lower Bound (CRLB) [25]. We generate 500 windows of size 7×7 with one spectral band having ensquared energy 0.5. One object is located at the center of the observed windows ($x_1 = 0, y_1 = 0$), and the second object is located 0.7 pixels away along the x-axis ($x_2 = 0.7, y_2 = 0$). Results are shown for the object within the central pixel as they are similar for both objects. MSE for x_{e_1} and y_{e_1} are similar. SPLEOv2 is a biased estimator because the training data is biased. A biased estimator can perform better than CRLB, as observed in Fig.

11. Since SPLEOv2 is biased, Reagan et al. [26] with SPLEOv2 as an initial estimate is also biased, outperforming CRLB. Performance is similar to CRLB at high SNRs as the localization becomes easier, and CRLB gets close to 0. SPLEOv2 provides a bad amplitude estimate with only one spectral band and does not achieve CRLB.

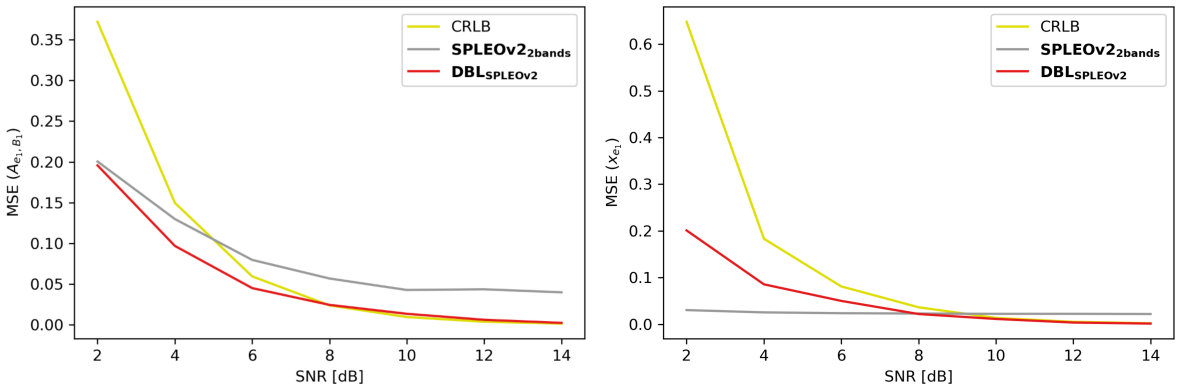

Fig. 12: Comparison of mean squared error in parameter estimates obtained from SPLEOv2 with Cramér-Rao Lower Bound. One object is at the center of the observed window. The second object is 0.7 pixels away along the x-axis. Results are shown for the first object. The observed window has 2 spectral bands.

Fig. 12 shows a comparison of mean squared error (MSE) in the parameter estimates (x_{e_1} and A_{e_1,B_1}) obtained from SPLEOv2 with the Cramér-Rao Lower Bound (CRLB) [25]. We generate 500 windows of size 7×7 with ensquared energy 0.5 for both spectral bands. One object is located at the center of the observed windows ($x_1 = 0, y_1 = 0$), and the second object is located 0.7 pixels away along the x-axis ($x_2 = 0.7, y_2 = 0$). Results are shown for the object within the central pixel as they are similar for both objects. MSE for x_{e_1} and y_{e_1} are similar and MSE for A_{e_1,B_1} and A_{e_1,B_2} are similar. We generate 500 windows of size 7×7 with ensquared energy 0.5 for both spectral bands. Similar to the single band case, SPLEOv2 trained on two spectral bands is also a biased estimator because the training data is biased. A biased estimator can perform better than CRLB, as observed in Fig. 12. Since SPLEOv2 is biased, DBL with SPLEOv2 as an initial estimate is also biased, outperforming CRLB. Performance is similar to CRLB at high SNRs as the localization becomes easier, and CRLB gets close to 0. SPLEOv2 provides a good amplitude estimate when trained on two spectral bands and achieves CRLB for SNR $> 4dB$.

5. CONCLUSIONS

In this paper, we present a machine learning object localization method, SPLEOv2, that works well for objects at low SNR and can also be used as an initial estimate for existing methods. SPLEOv2 can also be deployed on systems like satellites as it employs a lightweight CNN. Its low computation time also allows it to be used as an initial estimate for existing methods. SPLEOv2 has a low standard deviation and 95^{th} percentile localization error, meaning it performs consistently well for different observed windows at a particular SNR and ensquared energy. SPLEOv2's amplitudes estimates are not good but can be refined by using them as an initial estimate for other methods. Future work will explore training the network on more data to improve the amplitude estimate. It will also explore training on unbiased data to develop an unbiased estimator.

6. ACKNOWLEDGEMENTS

This material is based on research sponsored by Lockheed Martin Corporation. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied of Lockheed Martin Corporation.

Address all correspondence to Edward Delp, ace@ecn.purdue.edu

REFERENCES

[1] Anish V Abraham, Sripad Ram, Jerry Chao, ES Ward, and Raimund J Ober. Quantitative study of single molecule location estimation techniques. *Optics express*, 17(26):23352–23373, December 2009.

[2] Stephen M Anthony, John Mulcahy-Stanislawczyk, Eric A Shields, and Drew P Woodbury. Distinguishing one from many using super-resolution compressive sensing. *Proceedings of the Compressive Sensing VII: From Diverse Modalities to Big Data Analytics*, 10658:106580F, May 2018. Orlando, Florida.

[3] Hazen P Babcock, Jeffrey R Moffitt, Yunlong Cao, and Xiaowei Zhuang. Fast compressed sensing analysis for super-resolution imaging using l1-homotopy. *Optics express*, 21(23):28583–28596, November 2013.

[4] Emmanuel Bertin and Stephane Arnouts. Sextractor: Software for source extraction. *Astronomy and astrophysics supplement series*, 117(2):393–404, June 1996.

[5] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, September 2006.

[6] Gong Cheng, Xiang Yuan, Xiwen Yao, Kebing Yan, Qinghua Zeng, and Junwei Han. Towards large-scale small object detection: Survey and benchmarks. *arXiv preprint arXiv:2207.14096*, 2022.

[7] Haskell B Curry. The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261, 1944.

[8] Johan de Rooi and Paul Eilers. Deconvolution of pulse trains with the l0 penalty. *Analytica chimica acta*, 705(1-2):218–226, October 2011.

[9] Jinming Du, Huanzhang Lu, Moufa Hu, Luping Zhang, and Xinglin Shen. Cnn-based infrared dim small target detection algorithm using target-oriented shallow-deep features and effective small anchor. *IET Image Processing*, 15(1):1–15, January 2021.

[10] Yajuan Du, Hao Zhang, Mengying Zhao, Deqing Zou, and Chun Jason Xue. Faster super-resolution imaging of high density molecules via a cascading algorithm based on compressed sensing. *Optics express*, 23(14):18563–18576, July 2015.

[11] AM Ghez, Russel J White, and M Simon. High spatial resolution imaging of pre-main-sequence binary stars: resolving the relationship between disks and close companions. *The Astrophysical Journal*, 490(1):353, November 1997.

[12] Mridul Gupta, Sriram Baireddy, Jonathan Chan, Mitchell Krouss, Greg Furlich, Paul Martens, Moses W Chan, Mary L Comer, and Edward J Delp. Sub-pixel localization of objects using multiple spectral bands. *Proceedings of the IEEE Aerospace Conference*, pages 1–15, March 2022. Big Sky, MT.

[13] Mridul Gupta, Sriram Baireddy, Mary L Comer, Edward J Delp, Jonathan Chan, Mitchell Krouss, Paul Martens, and Moses Chan. Small target detection using optical flow. *Proceedings of the IEEE Aerospace Conference*, pages 1–9, March 2021. Big Sky, MT.

[14] Mridul Gupta, Jonathan Chan, Mitchell Krouss, Greg Furlich, Paul Martens, Moses W Chan, Mary L Comer, and Edward J Delp. Infrared small target detection enhancement using a lightweight convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 19, September 2022.

[15] Mridul Gupta, Jonathan Chan, Mitchell Krouss, Paul Martens, Moses W Chan, Mary L Comer, and Edward J Delp. Object localization in the presence of noise. *Proceedings of the IEEE Aerospace Conference*, pages 1–12, March 2023. Big Sky, MT.

[16] Samuel T Hess, Thanu PK Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, December 2006.

[17] Bo Huang, Hazen Babcock, and Xiaowei Zhuang. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell*, 143(7):1047–1058, December 2010.

[18] Siewert Hugelier, Johan J De Rooi, Romain Bernex, Sam Duwé, Olivier Devos, Michel Sliwa, Peter Dedecker, Paul HC Eilers, and Cyril Ruckebusch. Sparse deconvolution of high-density super-resolution images. *Scientific reports*, 6(1):1–11, February 2016.

[19] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[20] Faraz Lotfi, Farnoosh Faraji, and Hamid D Taghirad. Object localization through a single multiple-model switching cnn and a superpixel training approach. *Applied Soft Computing*, 115:108166, January 2022.

[21] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–

313, January 1965.
[22] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018.
[23] Donald M Olsson and Lloyd S Nelson. The nelder-mead simplex procedure for function minimization. *Technometrics*, 17(1):45–51, February 1975.
[24] Meenu Rani, Sanjay B Dhok, and Raghavendra B Deshmukh. A systematic review of compressive sensing: Concepts, implementations and applications. *IEEE Access*, 6:4875–4894, January 2018.
[25] C Radhakrishna Rao. Minimum variance and the estimation of several parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43(2):280–283, April 1947.
[26] John T Reagan and Theagenis J Abatzoglou. Model-based superresolution cso processing. *Proceedings of the Signal and Data Processing of Small Targets 1993*, 1954:204–218, October 1993.
[27] John T Reagan, Theagenis J Abatzoglou, John A Saghri, and Andrew G Tescher. Subpixel resolution for target tracking. *Proceedings of the Applications of Digital Image Processing XV*, 1771:2–19, January 1993.
[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, October 2015. Munich, Germany.
[29] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*, 3(10):793–796, October 2006.
[30] Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser. Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature methods*, 12(8):717–724, August 2015.
[31] Ankit Kumar Sharma and Hassan Foroosh. Slim-cnn: A light-weight cnn for face attribute prediction. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 329–335, November 2020. Buenos Aires, Argentina.
[32] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied optics*, 46(10):1819–1829, April 2007.
[33] Yan Zhou, Shaochang Chen, Yiming Wang, and Wenming Huan. Review of research on lightweight convolutional neural networks. *Proceedings of the IEEE Information Technology and Mechatronics Engineering Conference*, pages 1713–1720, June 2020. Chongqing, China.
[34] Lei Zhu, Wei Zhang, Daniel Elnatan, and Bo Huang. Faster storm using compressed sensing. *Nature methods*, 9(7):721–723, July 2012.