2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-2923-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISBI52829.2022.9761509

DEEP NEURAL NETWORK FOR 3D PARTICLE DETECTION IN 3D FLUORESCENCE MICROSCOPY IMAGES VIA DENSITY MAP REGRESSION

R. Spilger[1], *V. O. Chagin*[2,3], *C. S. Bold*[4], *L. Schermelleh*[5], *U. C. Müller*[4], *M. C. Cardoso*[2], *K. Rohr*[1]

[1]Biomedical Computer Vision Group, BioQuant, IPMB, Heidelberg University
[2]Cell Biology and Epigenetics, Department of Biology, Technische Universität Darmstadt
[3]Institute of Cytology, Russian Academy of Sciences, St. Petersburg
[4]Functional Genomics, IPMB, Heidelberg University
[5]Micron Advanced Bioimaging Unit, Department of Biochemistry, University of Oxford

ABSTRACT

Automatic detection of particles in fluorescence microscopy images is crucial to analyze cellular processes. We introduce a novel deep learning method for 3D fluorescent particle detection. Instead of pixel-wise binary classification or direct coordinate regression, we perform image-to-image mapping based on regressing a density map. Detections close to particles are rewarded in the network training, and highly nonlinear direct prediction of point coordinates is avoided. To focus on particles in comparison to background image points, we suggest using the adaptive wing loss. We also employ a weighted loss map to cope with the very strong imbalance between particle and background image points for 3D images. We evaluated our approach using 3D images of the Particle Tracking Challenge and real 3D fluorescence microscopy images of chromatin structures and interneurons. It turned out that our approach generally outperforms previous methods.

Index Terms— Biomedical imaging, fluorescence microscopy, 3D image data, particle detection, deep learning

1. INTRODUCTION

Particle detection in fluorescence microscopy images is a prerequisite for tracking and crucial to study cellular processes. Since manual detection of numerous particles is not feasible, fully automated approaches are required. Main challenges are low signal-to-noise ratio (SNR), small object size, clustering particles, and lack of appearance characteristics.

In previous work on fluorescent particle detection, classical methods were introduced (e.g., [1]) such as a wavelet-based detector [2], the spot-enhancing filter (SEF, [3]), a HDome transform-based detector [4], and adaptive thresholding with autoselected scale [5]. However, these methods are generally based on a predefined, relatively simple appearance model (e.g., Gaussian function), which does not necessarily hold. Recently, convolutional neural networks (CNNs) for particle detection have been introduced which can cope with more general appearance structures and show promising results (e.g., [6–9]). In [6–8], image-to-image mapping is performed based on pixel-wise binary classification, where each particle is represented by one or a few pixels in the binary ground truth mask. However, detections close to a particle but outside the particle region in the binary ground truth mask are not rewarded during network training. This decreases the stability of network training and makes training more difficult. In addition, [6] use a network with a relatively large number of parameters, and [7] employ a sliding window scheme which increases the computational cost. [9] use a CNN to directly regress offsets of bounding boxes, which, however, involves a highly nonlinear mapping from input images to point coordinates. None of the previous deep learning methods employs density map regression for particle detection, which is often used for key point detection in videos of natural scenes (e.g., faces, persons) yielding state-of-the-art performance (e.g., [10, 11]). In [12], density map regression is employed for cell counting in 2D images. A mean square error loss is used, which is not sensitive to small errors and not adaptive. Moreover, all deep learning methods above perform 2D detection using 2D CNNs, thus valuable volumetric information of 3D microscopy images is not exploited.

In this contribution, we introduce a novel deep learning approach for 3D particle detection in 3D fluorescence microscopy images. Instead of pixel-wise binary classification [6–8] or direct coordinate regression [9], we perform image-to-image mapping based on regressing a density map. The density map encodes the probability that a particle is located at a certain position. Thus, highly nonlinear direct prediction of point coordinates is avoided, and detections close to particles are rewarded in the network training. In addition, compared to [6–9], we exploit uncertainties in the manually annotated ground truth positions of particles for network training. To focus on particles in comparison to background image points, we suggest using the adaptive wing loss which was previously used for face recognition [10]. To cope with the

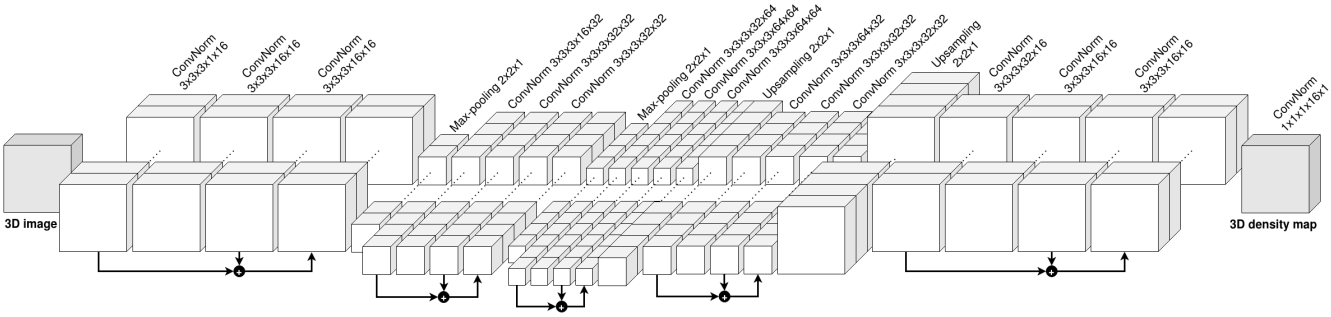

Fig. 1: Architecture of the proposed DM-DetNet3D network.

(a) Original image (b) Pixel-wise binary classification (c) Density map regression

Fig. 2: Original image and different types of ground truth. For visualization, all z-slices are max-pooled into one slice.

very strong imbalance between particle and background image points for 3D images, we use a weighted loss map, which assigns high weights to particles and difficult background image points close to particles. Compared to [10], where separate density maps are computed for each key point, in our approach all particles of an image are represented by only one density map. Different to [7], a sliding window scheme is not required, and all particles within an image are detected at once by sharing full-image convolutional features. Our method is the first that performs image-to-image mapping via density map regression for particle detection in microscopy images. In contrast to [6–12], the full 3D image information is exploited. The proposed 3D particle detection approach has been evaluated using 3D data of the Particle Tracking Challenge (PTC, [13]) as well as real 3D fluorescence microscopy images of chromatin structures and interneurons. It turned out that our approach outperforms previous methods.

2. METHODS

Our proposed deep learning approach for 3D particle detection, denoted as Density Map DetNet 3D (DM-DetNet3D), performs image-to-image mapping via density map regression. An overview of the network architecture is given in Fig. 1. The network is based on the slim hourglass architecture of DetNet [8], which was used for 2D images. The network is composed of a contracting and expanding path, can handle objects at different scales, and does not require a sliding window scheme. To detect all particles within an image at once, full-image convolution features are shared. Since detailed boundary information is not needed to detect sub-resolution particles and to reduce the number of parameters, long range skipping connections are not employed. To cope with the vanishing gradient problem, residual blocks [14] are used. Since batch normalization requires a representative data set to compute meaningful statistics, which is difficult to achieve when using only a few training samples, instance normalization is utilized. Compared to [8], we exploit the full 3D information by 3D convolution operations instead of 2D convolution operations in the residual blocks.

In contrast to pixel-wise binary classification [6–8] and direct coordinate regression [9], DM-DetNet3D performs image-to-image mapping based on regressing a density map which encodes the probability that a particle is located at a certain position. Thus, particle detections close to the correct position are taken into account in the network training, and highly nonlinear direct prediction of point coordinates is avoided. In our method, ground truth points in the training data are treated as Gaussian distributions centered around the annotated positions rather than using discrete image points as in [8]. This reflects that manual annotation of noisy, sub-resolution particles is generally uncertain, particularly for 3D images. The level of uncertainty is represented by the standard deviations $\sigma_{x,y}$ and σ_z of the Gaussian distribution. To exploit the full range of the sigmoid function in the output layer of the network, we normalize the values of the ground truth density maps to the range $[0, 1]$. During inference, the network predicts a density map from which particle positions are obtained by determining local maxima. Fig. 2 illustrates the type of ground truth for particle detection by density map regression compared to pixel-wise binary classification.

For accurate particle detection and localization via density map regression, the prediction accuracy of the network for image points representing particles is very important, since even small errors have a large effect. In comparison, for background image points the prediction accuracy is less important, since small errors usually have a small effect. Thus, the network should focus on particle image points during training (increased sensitivity). This can be achieved by using the adaptive wing loss (AWing, [10]), which adapts to dif-

ferent values in the ground truth mask to increase the sensitivity to errors for particles compared to background image points. This is an advantage over the standard mean square error (MSE) loss, which is insensitive to small errors and not adaptive causing blurred and dilated density maps. AWing is an extension of the wing loss for density map regression and is differentiable around zero. For a 3D image with width w, height h, and depth d, the AWing loss for the position x_i in the predicted density map $\boldsymbol{X} \in [0,1]^{h\times w\times d}$ is defined by:

$$\text{AWing}(x_i, y_i) = \begin{cases} \omega \ln(1 + |\frac{y_i - x_i}{\epsilon}|^{\alpha - y_i}) & \text{if } |y_i - x_i| < \theta \\ A|y_i - x_i| - C & \text{otherwise} \end{cases} \tag{1}$$

where y_i is the position in the ground truth density map $\boldsymbol{Y} \in [0,1]^{h\times w\times d}$. The parameters $A = \omega(1/(1+(\theta/\epsilon)^{(\alpha-y_i)}))(\alpha - y_i)((\theta/\epsilon)^{(\alpha-y_i-1)})(1/\epsilon)$ and $C = (\theta A - \omega \ln(1+(\theta/\epsilon)^{(\alpha-y_i)}))$ ensure that the loss is continuous and smooth at $|y_i - x_i| = \theta$. $\theta \in [0,1]$ is a threshold for switching between the linear and nonlinear part, $\alpha - y_i$ is used to adapt the curvature of the loss function to y_i (α has to be slightly larger than two). $\epsilon \in \mathbb{R}_{>0}$ limits the curvature of the nonlinear part and should not be set to a very small value since this would cause unstable training and exploding gradients for very small errors. In our experiments, we used $\alpha = 2.1$, $\omega = 14$, $\epsilon = 1$, and $\theta = 0.5$. Computing $\text{AWing}(x_i, y_i)$ for all positions in the predicted density map defines the *AWing loss map* $\mathcal{L}(\boldsymbol{X}, \boldsymbol{Y})$. In contrast to [10], we employ AWing to predict Gaussian distributions for all particles in an image using a *single* density map, whereas there for each key point a separate density map was used. Also, there a different application (face recognition) and 2D images were considered. We also tested our network using an MSE loss, which did not yield good results.

To address the very strong imbalance between particle and background image points for 3D images, we also use a *weight map* $\boldsymbol{W}$, which assigns high weights to particles and difficult background image points close to particles compared to other background image points:

$$\boldsymbol{W} = \begin{cases} \lambda + 1 & \text{for } \boldsymbol{Y}^d \geq T \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

where $\boldsymbol{Y}^d \in [0,1]^{h\times w\times d}$ is obtained by a 3×3 dilation of $\boldsymbol{Y}$ [10], λ defines the strength of the weighting (we used $\lambda = 10$), and $T \in [0,1]$ is a threshold. In $\boldsymbol{W}$, $\lambda + 1$ is assigned to particle image points and background image points close to particles, and 1 to all other image points. The *weighted loss map* $\mathcal{L}_\text{w}(\boldsymbol{X}, \boldsymbol{Y})$ between $\boldsymbol{X}$ and $\boldsymbol{Y}$ combines the weight map $\boldsymbol{W}$ and the AWing loss map $\mathcal{L}(\boldsymbol{X}, \boldsymbol{Y})$ and is defined by:

$$\mathcal{L}_\text{w}(\boldsymbol{X}, \boldsymbol{Y}) = \mathcal{L}(\boldsymbol{X}, \boldsymbol{Y}) \otimes \boldsymbol{W} \tag{3}$$

where $\otimes$ denotes the Hadamard product.

During network training, the mean value of $\mathcal{L}_\text{w}(\boldsymbol{X}, \boldsymbol{Y})$ is minimized using the AMSGrad optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to 0.001 and a mini-batch size of 4 was used. Data augmentation involved random cropping as well as horizontal and vertical flipping. To avoid overfitting, we applied early stopping after convergence is reached. The data set was randomly split into 50% for training, 25% for validation, and 25% for testing.

3. EXPERIMENTAL RESULTS

3.1. Particle Tracking Challenge Data

We evaluated DM-DetNet3D using 3D images of the PTC [13] and performed a comparison with the 3D versions of SEF [3] (SEF3D) and DetNet [8] (DetNet3D). SEF3D is based on the Laplacian-of-Gaussian, and DetNet3D performs image-to-image mapping by voxel-wise binary classification.

We used all 3D images of the virus scenario from the PTC comprising different object densities (low, medium, high) and SNR levels (SNR $= 1, 2, 4, 7$). In total, we considered 1200 images ($512 \times 512 \times 10$ voxels) and studied the results for each SNR level (300 images per SNR level). To measure the detection performance, we computed the F1 score $\in [0,1]$ using a gate of 5 voxels. The localization accuracy of correct particle detections is measured by the root mean square error (RMSE). For each SNR level, we averaged the F1 score and the RMSE over the respective images. The results for all SNR levels as well as the average values over the SNR levels and corresponding standard deviations are provided in Table 1. For the F1 score, DM-DetNet3D outperforms the other methods for all SNR levels, especially for low SNR levels. For RMSE, DM-DetNet3D yields the best result in three out of four cases.

3.2. Real Fluorescence Microscopy Data

We also evaluated DM-DetNet3D based on real 3D live cell fluorescence microscopy images of chromatin structures acquired with super-resolution 3D structured illumination microscopy [15]. The data set comprises 60 3D images from five different temporal image sequences ($512 \times 512 \times 5$ voxels). Ground truth was determined manually for difficult image regions (2637 particle positions in total). Main challenges of the data set are clustering particles and varying SNR levels due to photobleaching over time. The performance values in Table 2 show that DM-DetNet3D yields the best result. Example results in Fig. 3 demonstrate that the result of DM-DetNet3D agrees well with the ground truth.

In addition, we used challenging real 3D confocal fluorescence microscopy images of calretinin-immunostained interneurons in hippocampus sections of mice. For this data set we used one image with 116 ground truth positions for training and validation, and one image with 196 ground truth positions for testing. DM-DetNet3D yields an F1 score of 0.747 and outperforms SEF3D and DetNet3D with F1 scores of 0.686 and 0.543, respectively. For RMSE, DM-DetNet3D

Table 1: Results for 3D images of the Particle Tracking Challenge (mean $\pm$ standard deviation).

	SEF3D		DetNet3D		DM-DetNet3D	
	F1	RMSE	F1	RMSE	F1	RMSE
SNR 1	0.376 ± 0.141	2.491 ± 0.157	0.544 ± 0.118	$\mathbf{1.363} \pm 0.132$	$\mathbf{0.623} \pm 0.116$	1.396 ± 0.119
SNR 2	0.891 ± 0.059	0.967 ± 0.046	0.896 ± 0.059	0.686 ± 0.034	$\mathbf{0.973} \pm 0.013$	$\mathbf{0.612} \pm 0.040$
SNR 4	0.993 ± 0.005	0.655 ± 0.019	0.970 ± 0.012	0.522 ± 0.027	$\mathbf{0.994} \pm 0.005$	$\mathbf{0.496} \pm 0.016$
SNR 7	0.994 ± 0.005	0.574 ± 0.011	0.982 ± 0.007	0.508 ± 0.012	$\mathbf{0.995} \pm 0.004$	$\mathbf{0.505} \pm 0.015$
Average	0.813 ± 0.295	1.172 ± 0.895	0.848 ± 0.206	0.770 ± 0.404	$\mathbf{0.896} \pm 0.182$	$\mathbf{0.752} \pm 0.432$

Table 2: Results for real 3D images of chromatin structures.

Method	F1	RMSE
SEF3D	0.774 ± 0.078	1.934 ± 0.243
DetNet3D	0.711 ± 0.068	1.817 ± 0.193
DM-DetNet3D	$\mathbf{0.820} \pm 0.035$	$\mathbf{1.773} \pm 0.214$

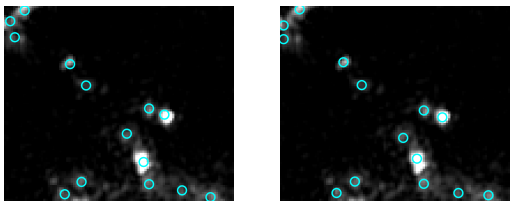

Fig. 3: Ground truth (left) and results of DM-DetNet3D (right) for a real 3D image of chromatin structures (section, maximum intensity projection).

(2.943) achieves similar results as SEF3D (2.871) and DetNet3D (2.835).

4. CONCLUSION

We presented a new deep learning approach for 3D particle detection in 3D fluorescence microscopy images that performs image-to-image mapping based on regressing a density map. During network training, detections close to particles are rewarded and uncertainties in the manually annotated ground truth positions are exploited. To focus on particles in comparison to background image points, we suggest using the adaptive wing loss. We also employ a weighted loss map to cope with the very strong imbalance between particle and background image points for 3D images. Our experiments for 3D images of the PTC and real 3D microscopy images show that our approach outperforms previous methods.

Acknowledgments: Support of the German Research Foundation (DFG) within the SPP 2202 (RO 2471/10-1, CA 198/15-1) and the SFB 1129 (project Z4), project number 240245660, is gratefully acknowledged.

Compliance with Ethical Standards: This study was performed retrospectively using animal subject data from a previous study conducted in line with the principles of the German Animal Welfare Act and approved by the Regierungspräsidium Karlsruhe (file number 35-9185.81/ G-151/15).

5. REFERENCES

[1] Štěpka et al., "Performance and sensitivity evaluation of 3D spot detection methods in confocal microscopy," *Cytometry A*, vol. 87, no. 8, pp. 759–772, 2015.

[2] J.-C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognit.*, vol. 35, no. 9, pp. 1989–1996, 2002.

[3] D. Sage et al., "Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics," *IEEE TIP*, vol. 14, no. 9, pp. 1372–1383, 2005.

[4] I. Smal et al., "A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering," in *Proc. IEEE ISBI*, 2008, pp. 264–267.

[5] A. Basset et al., "Adaptive spot detection with optimal scale selection in fluorescence microscopy images," *IEEE TIP*, vol. 24, no. 11, pp. 4512–4527, 2015.

[6] P. R. Gudla et al., "SpotLearn: Convolutional neural network for detection of fluorescence in situ hybridization (FISH) signals in high-throughput imaging approaches," in *Proc. CSH Symposia on Quant. Biol.*, 2017, pp. 57–70.

[7] J. M. Newby et al., "Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D," *PNAS USA*, vol. 115, no. 36, pp. 9026–9031, 2018.

[8] T. Wollmann et al., "DetNet: Deep neural network for particle detection in fluorescence microscopy images," in *Proc. IEEE ISBI*, 2019, pp. 517–520.

[9] F. Zakrzewski et al., "Automated detection of the HER2 gene amplification status in fluorescence in situ hybridization images for the diagnostics of cancer tissues," *Sci. Rep.*, vol. 9, no. 1, pp. 8231, 2019.

[10] X. Wang, L. Bo, and L. Fuxin, "Adaptive wing loss for robust face alignment via heatmap regression," in *Proc. IEEE/CVF ICCV*, 2019, pp. 6970–6980.

[11] Z. Luo et al., "Rethinking the heatmap regression for bottom-up human pose estimation," in *Proc. IEEE/CVF ICCV*, Jun. 2021, pp. 13264–13273.

[12] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," *Comput. Methods Biomech. Biomed. Eng.: Imaging Vis.*, vol. 6, no. 3, pp. 283–292, 2018.

[13] N. Chenouard et al., "Objective comparison of particle tracking methods," *Nat. Methods*, vol. 11, no. 3, pp. 281–289, 2014.

[14] K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE/CVF CVPR*, 2016, pp. 770–778.

[15] V. O. Chagin et al., "4D Visualization of replication foci in mammalian cells corresponding to individual replicons," *Nat. Commun.*, vol. 7, pp. 11231, 2016.