3D localization and autofocus of the particle field based on deep learning and depth-from-defocus

Zhao Dong
School of Mathematics and Physics
Hebei University of Engineering
dongzhao@hebeu.edu.cn

Shaokai Yang
Department of Physics
University of Alberta
shaokai1@ualberta.ca

Yan Sha
University of Alberta
ysha3@ualberta.ca

Abstract

Accurate three-dimensional positioning of particles is a critical task in microscopic particle research, with one of the main challenges being the measurement of particle depths. We present a novel approach for precise three-dimensional (3D) localization and autofocus of microscopic particles by integrating Depth-from-Defocus (DfD) techniques with deep learning. Our method combines You Only Look Once (YOLO) for lateral position detection with Generative Adversarial Networks (GANs) for autofocus, providing an efficient, noise-resistant, and real-time solution. Validated on synthetic datasets, static particle fields, and dynamic scenarios, the method achieved 99.9% accuracy on synthetic datasets and performed robustly on polystyrene particles, red blood cells, and plankton. Our algorithm can process a single multi-target image in 0.008 seconds, enabling real-time applications. Future work includes integrating Diffusion Models and the latest version of YOLO to enhance depth estimation and detection accuracy. Additionally, we are developing a user-friendly pipeline equipped with a graphical user interface (GUI) to make these advanced tools accessible to researchers across different disciplines, even those without prior deep learning expertise. This evolving pipeline will be continuously updated to improve precision and efficiency, making it a powerful and accessible tool for high-precision particle analysis in a wide range of scientific applications.

1 Introduction & Related Work

Particle field positioning is a crucial aspect across various domains, such as biomedical sciences, material science, and environmental engineering [1] [2]. In biomedical research, for example, precise three-dimensional localization of microscopic particles is essential for analyzing cellular behavior and developing drug delivery systems. Similarly, monitoring microparticles in water bodies is fundamental to environmental engineering. However, accurately determining the three-dimensional spatial information of particles, especially along the depth axis, remains a significant challenge. In practical applications, the lateral position of particles can often be obtained using centroid localization or object segmentation algorithms, but measuring the depth information is far more challenging [3] [4]. Researchers have proposed various approaches to overcome this difficulty, such as multi-particle imaging, accurate calibration for visual measurement, and digital holography [5-9]. Nevertheless, these methods face limitations in terms of adaptability to complex environments, high hardware requirements, and the trade-off between localization accuracy and computational efficiency. Depth-from-Defocus (DfD) is a technique that estimates depth by analyzing the extent of defocus in an image, originally proposed by Pentland in 1987 [10]. DfD has since been widely applied in depth measurement tasks, however, traditional DfD methods are prone to ambiguities, particularly when dealing with complex particle fields, resulting in unsatisfactory precision. [11-13] To improve localization accuracy, recent studies have incorporated deep learning into 3D localization tasks. Different Convolutional Neural Networks (CNNs) based models have been employed to enhance the

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

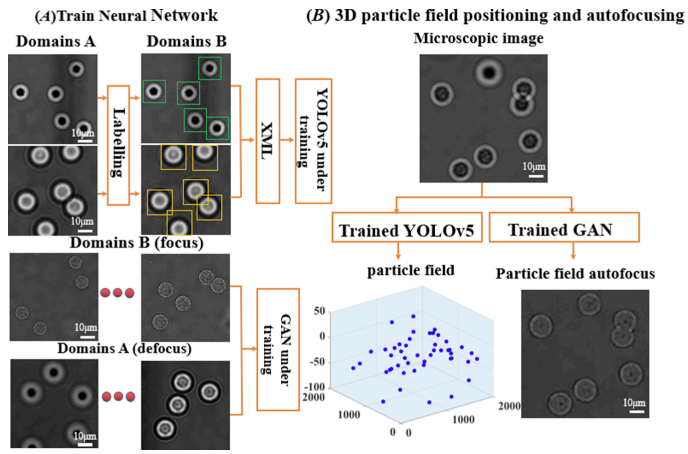

Figure 1: The workflow of the proposed method. (Step A) training neural network. (Step B) flow chart of the particle field positioning and autofocusing

precision of detecting blurred images [14-17], while Generative Adversarial Networks (GANs) have shown promise in autofocus tasks [18-25]. Against this backdrop, we propose a novel method for 3D localization and autofocus of particle fields based on DfD and deep learning. This approach combines the YOLO object detection network and GANs to achieve lateral position detection and depth-wise focusing of particles. Specifically, we utilize DfD to capture defocused images of particles, which are then processed by the YOLO version 5 (YOLOv5) network for automatic recognition of the 3D positions. Additionally, we employ Cycle-GAN and Pix2pix-GAN to achieve autofocus, resulting in clear particle imaging. Our method has demonstrated excellent results in accuracy and efficiency.

2 Methods

2.1 Workflow and Components of the Proposed Method

Our proposed method for precise 3D particle field localization, as shown in figure 1, combines the DfD with YOLOv5. During the training preparation stage, we used the annotation tool LabelImg [26] for preprocessing of the training images. Using LabelImg, we annotated the object categories and positional information within the images. In this process, depth was treated as the category name, and bounding boxes were manually drawn for each sample with a known depth. The lateral position was automatically obtained through these bounding boxes, and the relevant information was converted into XML files, which were used to prepare YOLOv5 for training. Additionally, we trained a GAN to obtain clear, focused images of particles. Defocused images were designated as Domain A, and focused images were designated as Domain B, which were used for effective GAN training. The trained YOLOv5 network outputs the 3D positions of particles, while the GAN, once well-trained, generates focused images of the particle fields. By seamlessly integrating YOLOv5 and GAN, our proposed method achieves accurate, efficient, and noise-resistant 3D localization and autofocus of particle fields, representing a significant advancement in this area.

2.2 Experimental Setup

Our experimental setup included capturing microscopic images of polystyrene particles using a commercial microscope (XSP-37XF, Shanghai Optical Instrument Factory, China). These particles were chosen due to their average diameter of approximately 10 microns and a refractive index of 1.587. In the experiments, a 40x objective lens was used to capture microscopic images of these particles, and a precision vertical translation stage was used to obtain images at different depths. During imaging, the particles were placed under Olympus immersion oil. Additionally, our method was tested on other particle fields, such as plankton and red blood cells, with more details about data acquisition available in [19].

Table 1. Comparisons of the training time and the parameters under different conditions.

Number of particles	Training		Parameters		
	Epochs	Time (h)	P	R	mAP
3361	200	0.946	0.748	0.944	0.905
	300	1.358	0.858	0.917	0.926
	600	2.758	0.864	0.935	0.905
5955	200	1.611	0.872	0.900	0.922
	300	**2.533**	**0.893**	**0.931**	**0.956**
	600	4.721	0.873	0.934	0.932
12 874	200	3.074	0.882	0.920	0.955
	300	4.599	0.888	0.921	0.955
	600	9.128	0.882	0.935	0.957

2.3 YOLO and GAN

YOLO is a real-time object detection system that has shown great potential for accurate detection of object positions. For our analysis, YOLOv5, the latest version of YOLO at the time, was used for its rapid detection and high precision. It employs a single-stage neural network to detect target positions directly. The model used in our study was a modified version of YOLOv5s, which has the smallest feature map depth and width in the YOLO series. Detailed network structures can be found in [27-30]. In our research, we employed two types of GANs to adapt to different conditions: Cycle-GAN and Pix2pix-GAN. Their structures and corresponding parameters can be found in [31].

2.4 Performance Evaluation

The trained model's ability to detect target particles was evaluated using precision (P), recall (R), average precision (AP), and mean average precision (mAP) [32]. Precision (P) is calculated as $P = \frac{TP}{TP+FP}$, where TP represents the number of correctly detected particles (True Positives), and FP represents the number of false detections (False Positives), measuring the accuracy of the model in detecting targets. Recall (R) is defined as $R = \frac{TP}{TP+FN}$, where FN represents the number of particles that were not detected (False Negatives), indicating the model's ability to detect all target instances. Average precision (AP) is computed by calculating the area under the precision-recall curve: $AP = \int_0^1 P(R)\,dR$, while mean average precision (mAP) is given by $mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$, providing an overall evaluation of the model's performance across all categories.

2.5 YOLOv5s Loss Function

The YOLOv5s model uses a loss function consisting of three components: bounding box regression loss (box-loss), classification loss (cls-loss), and objectness loss (obj-loss) [33]. During training, monitoring the loss curve helps determine whether the network model is converging steadily as iterations increase. The experimental results, shown in the figure 2's A section, show that the number of iterations increased, the loss values decreased when training and validating the model on polystyrene particles, indicating that the network achieved stable convergence, as shown in table 1.

3 Results of 3D Localization Method and Autofocus

3.1 Validation on Synthetic Dataset, Static, and Dynamic Scenarios

We first validated the proposed 3D localization method on a synthetic dataset. The synthetic dataset was generated using MicroSIG, a 3D ray-tracing-based synthetic image generator proposed by Rossi [34], to simulate the 3D distribution of particles. The trained YOLOv5 network effectively predicted the 3D positions of the particles, with depth information encoded through color coding. The experimental results showed, as illustrated in figure 2's B section, that YOLOv5 achieved an accuracy of 99.9% on the synthetic dataset, with most particle positions being accurately predicted, and only a few particles showing errors due to interference. As shown in figure 2's C section, in the static particle field, we applied the method to polystyrene particles and red blood cells, achieving validation accuracies of 99% and 97.8%, respectively. Even in the presence of overlapping particles or background noise, our method demonstrated high robustness, successfully detecting and localizing most particles. By capturing video of polystyrene particles moving in oil, we further verified the 3D localization capability of the trained YOLOv5 network in dynamic scenarios. The processing time

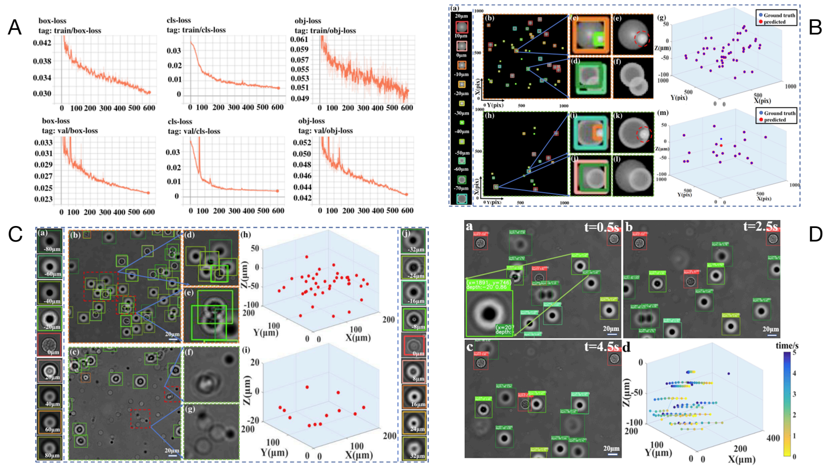

Figure 2: A. The variation curves of the loss values; B. Application of proposed method to the datasets generated by MicroSIF; C. Typical microscopic images of the particle field; D. The motions of particles in the chamber.

for each frame was approximately 0.008 seconds, meeting the requirements for real-time detection. Additionally, experimental results with plankton samples showed good detection performance, with all plankton samples successfully detected and accurately localized, as shown in figure 2's D section. The 3D movement trajectories of plankton provide data support for further behavioral research. It should be noted that in some cases, contaminants with similar color and morphology to plankton may be misidentified as plankton, resulting in detection errors.

3.2 Implementation and Performance Evaluation of Autofocus

To achieve particle autofocus, we used two types of Generative Adversarial Networks (GANs): Cycle-GAN and Pix2Pix-GAN. In static particle fields, due to the large amount of data available for polystyrene particles and red blood cells, Cycle-GAN performed well and was able to convert particle images at different depths into clear, focused images. However, for plankton samples, due to the limited amount of data, Cycle-GAN struggled to handle these samples effectively, and therefore, we used Pix2Pix-GAN. By employing data augmentation, we generated paired defocused and focused images and trained Pix2Pix-GAN to achieve autofocus for plankton. The experimental results showed that the Structural Similarity Index (SSIM) [35-37] between GAN-generated autofocus images and real images was approximately 0.95, confirming the feasibility and effectiveness of this method in complex scenarios.

4 Summary and Future Work

In this paper, we proposed a method for 3D particle localization and autofocus that combines YOLOv5 with GANs, validated on synthetic datasets, static, and dynamic particle fields. Experimental results demonstrated our method's accuracy, robustness, and excellent performance across these scenarios. By leveraging YOLOv5 for 3D position detection and GANs for autofocus, we achieved efficient, noise-resistant, and real-time 3D localization and focusing. However, challenges such as misidentification of contaminants and depth estimation errors for overlapping particles remain for future improvement. To enhance performance, we are integrating newer YOLO versions and Diffusion Models to improve depth estimation and detection accuracy. Diffusion Models can refine depth estimation and image quality, while the advanced YOLO framework aims to boost detection precision and efficiency. We plan to develop a user-friendly pipeline with a graphical user interface (UI) to make these tools accessible for researchers without deep learning expertise. The UI will enable users to run analyses and visualize results interactively, without needing in-depth technical knowledge. Continuous updates will further improve accuracy and efficiency, making it an evolving tool for particle field research.

References

[1] Wu, Y.C., Shiledar, A., Li, Y.C., Wong, J., Feng, S., Chen, X., and Ozcan, A. (2017). Air quality monitoring using mobile microscopy and machine learning. In *Light: Science & Applications*, 6, 17046. Cambridge, MA: Springer Nature.

[2] Yoon, J., Jo, Y., Kim, M.H., Kim, K., Lee, S., Kang, S.J., and Park, Y. (2017). Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning. In *Scientific Reports*, 7, 1-10. Cambridge, MA: Nature Publishing Group.

[3] Gao, Z., Wu, Y., and Bao, Y. (2018). Image analysis for in-line measurement of multidimensional size, shape, and polymorphic transformation of L-glutamic acid using deep learning-based image segmentation and classification. In *Crystal Growth & Design*, 18, 4275-4281. Cambridge, MA: American Chemical Society.

[4] Tsalicoglou, C., and Rösgen, T. (2022). Deep learning based instance segmentation of particle streaks and tufts. In *Measurement Science and Technology*, 33, 114005. Cambridge, MA: IOP Publishing.

[5] Cierpka, C., and Kähler, C.J. (2012). Particle imaging techniques for volumetric three-component (3D3C) velocity measurements in microfluidics. In *Journal of Visualization-Japan*, 15, 1-31. Cambridge, MA: Springer.

[6] Cui, Y., Zhou, F., Wang, Y., Liu, L., and Gao, H. (2014). Precise calibration of binocular vision system used for vision measurement. In *Optics Express*, 22, 9134-9149. Cambridge, MA: Optical Society of America.

[7] Dan, H.C., Bai, G.W., Zhu, Z.H., Liu, X., and Cao, W. (2022). An improved computation method for asphalt pavement texture depth based on multiocular vision 3D reconstruction technology. In *Construction and Building Materials*, 321, 126427. Cambridge, MA: Elsevier.

[8] Chen, N., Wang, C., and Heidrich, W. (2021). Holographic 3D particle imaging with model-based deep network. In *IEEE Transactions on Computational Imaging*, 7, 288-296. Cambridge, MA: IEEE.

[9] Lee, S.J., Yoon, G.Y., and Go, T. (2019). Deep learning-based accurate and rapid tracking of 3D positional information of microparticles using digital holographic microscopy. In *Experimental Fluids*, 60, 1-10. Cambridge, MA: Springer.

[10] Pentland, A.P. (1987). A new sense for depth of field. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4, 523-531. Cambridge, MA: IEEE.

[11] Zhou, W., Jin, N., Jia, M., Yang, H., and Cai, X. (2016). Three-dimensional positioning method for moving particles based on defocused imaging using single-lens dual-camera system. In *Chinese Optics Letters*, 14, 031201. Cambridge, MA: Chinese Laser Press.

[12] Barnkob, R., Cierpka, C., Chen, M., Sachs, S., Mäder, P., and Rossi, M. (2021). Defocus particle tracking: a comparison of methods based on model functions, cross-correlation, and neural networks. In *Measurement Science and Technology*, 32, 094011. Cambridge, MA: IOP Publishing.

[13] Rossi, M., and Barnkob, R. (2020). A fast and robust algorithm for general defocusing particle tracking. In *Measurement Science and Technology*, 32, 014001. Cambridge, MA: IOP Publishing.

[14] Franchini, S., and Krevor, S. (2020). Cut, overlap and locate: a deep learning approach for the 3D localization of particles in astigmatic optical setups. In *Experimental Fluids*, 61, 1-15. Cambridge, MA: Springer.

[15] Nehme, E., Freedman, D., Gordon, R., Ferdman, B., Weiss, L.E., Alalouf, O., and Shechtman, Y. (2020). DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. In *Nature Methods*, 17, 734-740. Cambridge, MA: Nature Publishing Group.

[16] Dreisbach, M., Leister, R., Probst, M., Friederich, P., Stroh, A., and Kriegseis, J. (2022). Particle detection by means of neural networks and synthetic training data refinement in defocusing particle tracking velocimetry. In *Measurement Science and Technology*, 33, 124001. Cambridge, MA: IOP Publishing.

[17] Sachs, S., Ratz, M., and Mäder, P. (2023). Particle detection and size recognition based on defocused particle images: a comparison of a deterministic algorithm and a deep neural network. In *Experimental Fluids*, 64, 21. Cambridge, MA: Springer.

[18] Leroy, R., Trouvé-Peloux, P., Le Saux, B., Buat, B., and Champagnat, F. (2022). Learning local depth regression from defocus blur by soft-assignment encoding. In *Applied Optics*, 61, 8843-8849. Cambridge, MA: Optical Society of America.

[19] Zhang, X., Wang, H., Wang, W., Yang, S., Wang, J., and Lei, J. (2022). Particle field positioning with a commercial microscope based on a developed CNN and the depth-from-defocus method. In *Optics & Laser Engineering*, 153, 106989. Cambridge, MA: Elsevier.

[20] Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., and Tian, X. (2022). An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease. In *Computers and Electronics in Agriculture*, 194, 106780. Cambridge, MA: Elsevier.

[21] Lei, F., Tang, F., and Li, S. (2022). Underwater target detection algorithm based on improved YOLOv5. In *Journal of Marine Science and Engineering*, 10, 310. Cambridge, MA: MDPI.

[22] Subbarao, M., and Tyan, J.K. (1998). Selecting the optimal focus measure for autofocusing and depth-from-focus. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 864-870. Cambridge, MA: IEEE.

[23] Akpinar, U., Sahin, E., and Gotchev, A. (2019). Learning optimal phase-coded aperture for depth of field extension. In *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4315-4319. Cambridge, MA: IEEE.

[24] Wu, Y., Rivenson, Y., Zhang, Y., Wei, Z., Günaydin, H., Lin, X., and Ozcan, A. (2018). Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. In *Optica*, 5, 704. Cambridge, MA: Optical Society of America.

[25] Zhu, J., Park, T., Isola, P., and Efros, A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242-2251. Cambridge, MA: IEEE.

[26] Liu, C. & Wu, Y. & Liu, J. & Sun, Z. (2021) Improved YOLOV3 network for insulator detection in aerial images with diverse background interference. In *Electronics*, 10, 771.

[27] Wang, L. & Yan, W. (2021) Tree leaves detection based on deep learning. In M. Nguyen, W. Yan, H. Ho (Eds.), *Geometry and Vision*, vol. 1386, pp. 26–38. Cham: Springer.

[28] Yan, B. & Fan, P. & Lei, X. & Liu, Z. & Yang, F. (2021) A real-time apple targets detection method for picking robot based on improved YOLOv5. In *Remote Sens.*, 13, 1619.

[29] Yao, J. & Qi, J. & Zhang, J. & Shao, H. & Yang, J. & Li, X. (2021) A real-time detection algorithm for kiwifruit defects based on YOLOv5. In *Electronics*, 10, 1711.

[30] Wu, D. & Lv, S. & Jiang, M. & Song, H. (2020) Using channel pruning-based YOLOv4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. In *Comput. Electron. Agric.*, 178, 105742.

[31] Wang, W. & Wang, H. & Yang, S. & Zhang, X. & Wang, X. & Wang, J. & Lei, J. & Zhang, Z. & Dong, Z. (2021) Resolution enhancement in microscopic imaging based on generative adversarial network with unpaired data. In *Opt. Commun.*, 503, 127454.

[32] Shinde, S. & Kothari, A. & Gupta, V. (2018) YOLO based human action recognition and localization. In *Proc. Comput. Sci.*, 133, pp. 831–8.

[33] Malta, A. & Mendes, M. & Farinha, T. (2021) Augmented reality maintenance assistant using YOLOv5. In *Appl. Sci.*, 11, 4758.

[34] Rossi, M. (2019) Synthetic image generator for defocusing and astigmatic PIV/PTV. In *Meas. Sci. Technol.*, 31, 017003.

[35] Tang, C. & Zhang, W. & Wang, L. & Cai, A. & Liang, N. & Li, L. & Yan, B. (2020) Generative adversarial network-based sinogram super-resolution for computed tomography imaging. In *Phys. Med. Biol.*, 65, 235006.

[36] Isola, J. & Zhu, T. & Zhou, A. & Efros, A. A. (2017) Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–76. Honolulu, HI, USA: IEEE.

[37] Wang, Z. & Bovik, A. C. & Sheikh, H. R. & Simoncelli, E. P. (2004) Image quality assessment: from error visibility to structural similarity. In *IEEE Trans. Image Process.*, 13, pp. 600–12.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes] ,

   Justification: The claims in the abstract and introduction accurately reflect the contributions of the paper, specifically highlighting the novel application of deep learning models in the analysis of biophysics data.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification:The paper includes a dedicated section on limitations (Section 4). We address the limitations related to the generalizability of our approach, as the results have been validated on a limited number of datasets.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This paper focuses on the application of deep learning models for 3D reconstruction and does not introduce any new theoretical results that require formal proofs. The methods and algorithms used are based on existing techniques, which are referenced appropriately.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides detailed descriptions of the experimental setup, data acquisition methods, and model architectures, including the use of super-resolution and diffusion models. Additionally, all code, model training scripts, and datasets have been made available on GitHub (as mentioned in Section 3.3), allowing for full reproducibility of the results. Furthermore, the paper outlines the specific hardware and software configurations used for training.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides open access to both the code and datasets through a publicly accessible GitHub repository, as stated in Section 3.3. Detailed instructions are included for data access, environment setup, and running the code, ensuring that the experiments can be faithfully reproduced.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies the details of the experimental setup, including data acquisition, data splits, and model hyperparameters. The choice of optimizer and its configurations are also provided, ensuring that the experimental results can be fully understood. Further details are provided in the supplemental material and the accompanying GitHub repository (Section 2.1 and 2.2)."

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The paper focuses on the qualitative performance of the Vision-Cell framework in sparse data environments and does not include statistical significance measures such as error bars. The experiments are designed to demonstrate the framework's ability to generate high-quality 3D reconstructions rather than to assess variability or uncertainty across multiple runs.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper specifies the type of compute resources used, including an RTX 4060 GPU and later two A100 GPUs for more complex model training. Memory limitations encountered with the RTX 4060 GPU are also discussed, along with the time required for training tasks. Further details, such as memory usage and time of execution, are provided in the supplemental material (see Section 4).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: he research conducted in this paper fully conforms with the NeurIPS Code of Ethics. All experiments and data used in this study are from publicly available sources or obtained through ethical means, with no human or animal subjects involved. The paper adheres to the principles of transparency, fairness, and societal responsibility outlined in the Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper discusses the positive societal impacts of the Vision-Cell framework, particularly in its potential to advance biological and medical research by enabling high-precision 3D cell reconstructions with minimal physical damage to samples. This could lead to breakthroughs in fields such as neuroscience and structural biology. The paper also considers the broader implications of deep learning in biological research but does not foresee any direct negative societal impacts, as the work is foundational and not tied to applications with ethical risks.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not release any data or models that pose a high risk for misuse. The Vision-Cell framework focuses on biological data with no foreseeable ethical concerns related to privacy, security, or malicious use. Therefore, no specific safeguards are necessary for the release of this work.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper uses publicly available datasets and models, all of which are properly credited in the references section. We have explicitly mentioned the license and terms of use for these assets where applicable (see Section X for details), and no scraped data from restricted sources were used

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The Vision-Cell framework introduced in this paper is a new asset, and detailed documentation is provided alongside the code and models, including instructions for usage, training, and evaluation. The assets are released via a GitHub repository with a structured template to ensure easy reproducibility (see Section 3.3).

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.

- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [No] ,

    Justification: The paper does not involve crowdsourcing or research with human subjects, as it focuses on computational experiments using biological data obtained through non-human means

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [No]

    Justification: The paper does not involve research with human subjects, and therefore, IRB approval or equivalent is not applicable.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.