Microplastic Identification Using AI-Driven Image Segmentation with Synthetic Ecological Context

Alex Dils[1], David Raymond[2], Jack Spottiswood[1], Samay Kodige[1]
Dylan Karmin[3], Rikhil Kokal[1], Win Cowger[4], Christoph Sadée[5]

[1] University of California, Berkeley
[2] Dartmouth College
[3] Massachusetts Institute of Technology
[4] Moore Institute for Plastic Pollution Research
[5] Division of Computational Medicine, Stanford University

Abstract

Conventional methods for microplastic identification in water samples are costly, slow, and dependent on specialized expertise. We present a deep learning segmentation framework for identifying microplastic foreground in microscopy images and evaluate whether synthetic ecological context can improve model performance when labeled real data are limited. We also contribute a curated image dataset with manually segmented microplastic masks, adding paired image-mask examples for supervised training and held-out ecological evaluation. The workflow combines manually labeled laboratory microplastic images with generated inpainting examples selected for visible foreground change, non-empty masks, plausible object scale, and background diversity. In our results, adding verified synthetic examples improves segmentation across architectures by increasing exposure to diverse ecological scenarios while preserving pixel-level labels. The strongest synthetic-assisted model reaches Dice 0.817, IoU 0.704, and boundary F1 0.858 on the held-out ecological evaluation set, compared with Dice 0.743, IoU 0.619, and boundary F1 0.809 for the strongest real-only model. These results extend prior microplastic segmentation work by testing synthetic ecological context under matched held-out evaluation and show that high-quality synthetic examples can improve segmentation.

1 Introduction

Microplastics are a widespread ecological concern, with detection and monitoring needs shaped by large-scale plastic production, environmental persistence, and heterogeneous particle types [1, 6, 7]. Microplastic monitoring needs methods that are accurate enough for scientific screening yet practical enough to apply at scale. Manual microscopy can flag candidate particles, but it is slow and subjective, and its reliability depends heavily on the analyst [2, 3, 7]. Spectroscopic techniques such as FTIR and Raman provide far more specific confirmation, but they are costly and low-throughput, which limits how widely they can be deployed [2, 3, 7]. Deep learning has already been explored for related microplastic analysis tasks, including spectral reconstruction and microscopy-based classification [9, 10]. A useful image-based system should sit between these extremes: it should help researchers prioritize candidate particles, quantify their morphology, and reduce the volume of manual review required before confirmatory analysis.

We approach this problem with semantic segmentation, a task with strong precedent in biomedical imaging [4, 18]. Segmentation would label microplastics at the pixel level, which makes it possible to recover the morphological quantities that matter for microplastic characterization: area, perimeter, width, shape, and contact with surrounding debris. Because a mask directly estimates which pixels belong to a particle, it is naturally suited to morphology and burden estimation. Detection remains valuable for triage and counting, but when the goal is quantitative measurement and morphology-aware screening, segmentation is the more appropriate first task.

The central obstacle to training such a model is the scarcity of labeled real-world microscopy images, particularly across diverse imaging conditions. Pixel-level labels are particularly limited in the ecological setting, even though that is where segmentation is hardest and most useful. Recent reviews also point to data standardization and sharing as recurring constraints for microplastics research [5, 8]. Synthetic inpainting offers a practical way to bridge this gap: it can place known particle masks into realistic ecological backgrounds, expanding the visual diversity of the training data while preserving ground-truth labels.

Recent microplastic segmentation studies demonstrate that the task is feasible: Park et al. developed MP-Net, a U-Net-derived model for fluorescence microscopy images of microplastics isolated from clams, and reported a mean F1-score of 0.736 and mean IoU of 0.617 [28]. Yao et al. reported 0.690 mIoU for a lightweight multi-class MNv4-Conv-M-fpn model [24], and Xu and Wang reported UNet and UNet2plus mIoUs of 91.45% and 91.08% on an urban-water microplastic dataset [25]. Those studies establish strong task-specific segmentation performance; the question here is whether synthetic ecological context improves transfer to held-out ecological microscopy under matched real-only and synthetic-assisted training conditions.

This motivates the central question of the study: does adding synthetic ecological context improve segmentation performance on real samples? To answer it rigorously, we evaluate the effect across several segmentation architectures, so that any observed benefit reflects a general property of the synthetic data rather than an artifact of a single model. All final claims rest on held-out ecological evaluation rather than on synthetic validation alone, since strong performance on synthetic backgrounds does not guarantee performance on the real samples the method is meant to analyze. The study therefore has two linked contributions: a manually segmented microplastic microscopy dataset that adds pixel-level image-mask labels, and a controlled benchmark testing whether synthetic ecological context improves segmentation on real held-out ecological samples.

2 Data

This study draws on three image cohorts, each serving a distinct role in training or evaluation. Together, the assembled dataset contains 1,198 microscopy images, including 465 manually segmented image-mask pairs and 733 unlabeled ecological background images.

Cohort 1 consists of microscopy images of laboratory-prepared microplastic samples derived from the Moore Institute for Plastic Pollution Research data resources. Each microplastic particle was manually segmented and converted into a binary mask, yielding a one-to-one correspondence between images and labels. This cohort serves two purposes: it supplies the labeled examples for real supervised training, and its masks provide the particle templates inserted into ecological backgrounds during inpainting.

Cohort	Source type	Images	Masks	Role in study
Cohort 1	Laboratory microplastic microscopy	368	368	Real supervised training source and microplastic mask source for inpainting
Cohort 2	Ecological microscopy without microplastic	733	0	Background image source for synthetic ecological context generation
Cohort 3 (test set)	Ecological microplastic microscopy	97	97	Held-out evaluation set for final segmentation metrics

Cohort 2 comprises ecological microscopy images that contain no annotated microplastics. Because these images capture realistic environmental clutter—sediment, bubbles, fibers, and organic matter—they serve as the background source into which known particle masks are inpainted, expanding the visual diversity of the training data while preserving ground-truth labels.

Cohort 3 is the held-out evaluation set and is never seen during training. It contains labeled microplastic images drawn from ecological settings that are visually distinct from those in the other cohorts, with each particle manually segmented into a binary mask. Reserving this cohort for final evaluation ensures that all reported metrics reflect performance on real ecological samples rather than on synthetic data. The manual segmentation in Cohorts 1 and 3 is a dataset contribution of the study, since these masks convert microscopy images into reusable pixel-level labels for training, benchmarking, and morphology-aware evaluation.

3 Models

3.1 Stable Diffusion Inpainting Model Training

The custom inpainting generator was initialized from the pretrained Stability AI Stable Diffusion 2 inpainting checkpoint and fine-tuned on Cohort 1 laboratory microplastic images with their dilated binary masks; the full checkpoint identifier is reported in the model-settings appendix. Each training item consisted of the microscopy image, its corresponding inpainting mask, and the masked-image conditioning input used by the Stable Diffusion inpainting formulation. Images and masks were resized to 512×512 pixels, and the model was trained to reconstruct the masked microplastic foreground while using the unmasked image region as visual context. The held-out Cohort 3 ecological evaluation images were not used in any inpainting-model training step. The fine-tuned component carried forward for generation was the Diffusers inpainting U-Net checkpoint. Training used seed 42, 100 epochs, batch size 1, gradient accumulation of 4, two data-loader workers, learning rate 0.00001, a constant learning-rate schedule with no warmup, AdamW optimization with $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay 0.01, epsilon 10^{-8}, maximum gradient norm 1.0, and fp16 mixed precision. Memory-efficient attention was enabled when available. Checkpoints were written every 500 steps and at the end of each epoch.

3.2 Synthetic Inpainting Generation

Before segmentation training, synthetic image/mask pairs were produced using generative inpainting [13, 14, 15, 16, 17] (Figures 1 and 2). Real Cohort 1 microplastic masks were geometrically transformed and inserted into unlabeled Cohort 2 ecological microscopy backgrounds to create masked ecological composites. The geometric transforms included random rotation, horizontal and vertical flipping, isotropic scaling, mild

anisotropic scaling, translation to a random valid background location, and shear. Each transformed mask was placed only where it remained inside the image frame, and the corresponding real microplastic crop was inserted into the ecological background to create a mask-guided editing composite. This mask-guided editing composite provided the spatial location and approximate particle structure for inpainting, while the binary transformed mask defined the region to be regenerated. Stable Diffusion inpainting [13, 14, 15] was then used to synthesize plausible microplastic foreground appearance inside the masked region while preserving the surrounding ecological background. For each candidate example, the input to the generator consisted of the ecological background, the mask-guided editing composite, the transformed binary mask, and a text prompt specifying a realistic microscope water sample with a visible microplastic fiber or fragment in the masked region. Negative prompts penalized cartoon-like output, text, watermarks, blur, unchanged backgrounds, and empty masks. The generation settings used 40–45 denoising steps, guidance scale 7.0–8.0, inpainting strength 0.99, full-mask diffusion, and 4-pixel mask dilation. The retained transformed mask served as the segmentation label for each generated image. Stable Diffusion inpainting was used to generate 10,000 candidate synthetic image/mask pairs. These generator outputs were not evaluated against the held-out Cohort 3 labels; they were used only to create candidate training pairs for the downstream segmentation experiment. During Stable Diffusion inpainting, generation was restricted to pixels inside the supplied binary mask, while all pixels outside the mask were preserved from the original ecological background, ensuring that only the masked foreground region was modified.

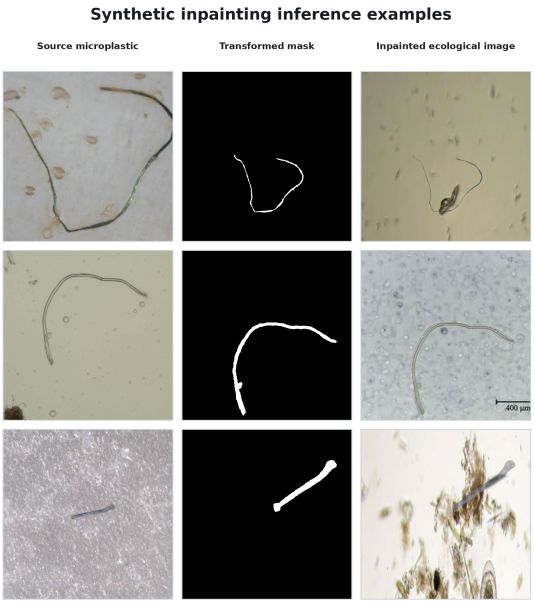

Figure 1. Three-by-three inference grid showing representative source microplastic examples with original masks, transformed masks, and inpainted ecological images.

3.3 Quality Control on Synthetic Inpainting

Synthetic inpainting examples were quality-controlled by scoring each generated image with four mask-region change metrics: masked mean absolute difference versus the original ecological background, changed-pixel fraction versus the original background, masked mean absolute difference versus the mask-guided edit-

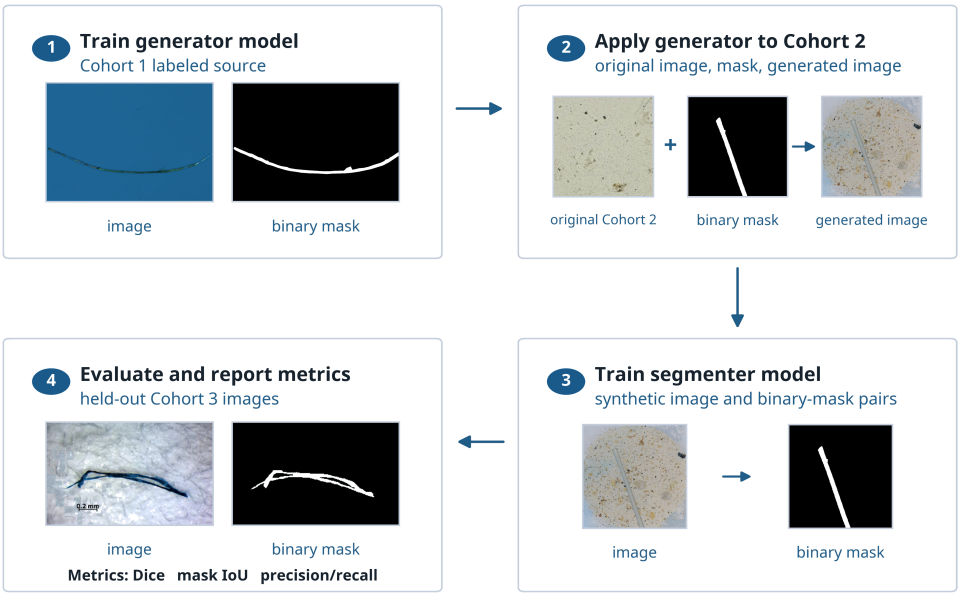

Figure 2. Synthetic ecological context generation and evaluation pipeline. A generator fine-tuned from Cohort 1 image-mask pairs is applied to Cohort 2 backgrounds, synthetic image-mask pairs train the segmenter, and performance is evaluated on held-out Cohort 3 images.

ing composite, and changed-pixel fraction versus the mask-guided editing composite. Foreground mask fraction was also checked to exclude empty, tiny, or implausibly large masks. Candidate images were ranked by a composite normalized QC score, and the final top-inpainting subset was selected with diversity caps of one image per ecological background and two images per source microplastic mask. This retained generated examples that visibly changed the masked region, preserved plausible microplastic scale, and avoided overrepresenting repeated backgrounds or source masks.

$$
\begin{aligned}
Q_i &= \sum_{k \in \mathcal{K}} z_{ik} - 0.25\, \frac{|m_i - \mathrm{median}(m)|}{q_{0.95,m} - q_{0.05,m}}, \\
z_{ik} &= \frac{\mathrm{clip}\left(x_{ik}, q_{0.01,k}, q_{0.99,k}\right) - q_{0.01,k}}{q_{0.99,k} - q_{0.01,k}}, \\
\mathcal{K} &= \{\text{masked MAD vs background, changed fraction vs background,} \\
&\qquad \text{masked MAD vs composite, changed fraction vs composite}\}, \\
m_i &= \text{mask foreground fraction.}
\end{aligned}
\tag{1}
$$

3.4 Segmentation Model Training

The full benchmark evaluates U-Net [18], U-Net++ [19], DeepLabV3+ [20], FPN [21], SegFormer-B2 [22], and YOLO11m-seg [23]. Models are trained with Dice loss [26] and binary cross-entropy. A fixed threshold of 0.5 means that pixels with predicted foreground probability at or above 0.5 are counted as microplastic foreground during evaluation. All models use deterministic seeds 13, 37, and 101; images are resized to 512×512 during training; batch size is 8; the maximum training horizon is 80 epochs; the learning rate is

0.0001 with weight decay 0.00001; early stopping patience is 15 epochs; and mixed precision is enabled when CUDA is available. Training augmentations include horizontal and vertical flips, translation, crop, shear, and color jitter. The primary metrics are Dice and IoU; secondary metrics include boundary F1 [27], precision, recall, and area error. Validation scores are used only for early stopping and diagnostics. The central evidence comes from the held-out ecological evaluation set.

4 Experimental Design

The experiment is designed to isolate the effect of synthetic ecological context on segmentation performance. Every condition uses the same real labeled images, the same evaluation set, the same segmentation threshold, and the same training schedule. The comparison changes only whether synthetic inpainting examples are added to the real training data and, in the quality-controlled condition, which generated images are admitted. Synthetic examples are generated by placing transformed real microplastic masks into ecological backgrounds and using inpainting to synthesize foreground appearance inside the masked region. The retained mask is used as the label. For the verified condition, the hard exclusions are non-empty masks and diversity caps of one image per ecological background and two images per source microplastic mask; the remaining candidates are ranked by the normalized QC score in Equation 1 rather than by a single manual cutoff. This prevents the model from learning from no-op generations, oversized artifacts, or repeated backgrounds.

Condition	**Training data**	**Purpose**
Real only	Labeled Cohort 1 images only	Baseline for supervised segmentation without synthetic ecological context
Real + unfiltered inpainting	Real images plus all generated inpainting examples	Tests whether synthetic scale alone is sufficient
Real + verified inpainting	Real images plus QC-filtered inpainting examples	Tests whether high-quality synthetic ecological context improves transfer
Half real + top inpainting pilot	Equal real and top-ranked inpainting examples	Tests whether a smaller set of best synthetic samples can outperform larger noisier synthetic sets

5 Results

Overall, quality-controlled inpainting improves held-out ecological segmentation across the retained model matrix. The strongest verified-inpainting checkpoint is U-Net++ at seed 37, with Dice 0.817, IoU 0.704, boundary F1 0.858, precision 0.802, and recall 0.846. The strongest real-only checkpoint reaches Dice 0.743, IoU 0.619, and boundary F1 0.809. Thus, the top synthetic-assisted run improves Dice by 0.074 and IoU by 0.085 over the strongest real-only run. As an external point of reference, MP-Net achieved mean F1 0.736 and mean IoU 0.617 on fluorescence microscopy images of microplastics isolated from clams [28]. Our strongest synthetic-assisted checkpoint is numerically higher on Dice/F1 and IoU, but this is not a direct benchmark comparison because the imaging modality, sample source, label protocol, and test distribution differ. The synthetic effect is not limited to a single architecture. All five retained semantic architectures improve under

the verified-inpainting condition. The strongest absolute performance remains with U-Net++ and SegFormer-B2, suggesting that synthetic ecological context is most effective when paired with architectures that already handle scale and boundary detail well.

Table 1: Held-out ecological segmentation metrics by training condition. Values are mean ± standard deviation where available; bold entries mark the best value in each metric column.

Training condition	**Runs**	**Dice**	**IoU**	**Boundary F1**	**Best validation Dice**
Real only	18	0.641 ± 0.161	0.519 ± 0.138	0.703	0.758
Real + unfiltered inpainting	18	0.586 ± 0.143	0.462 ± 0.132	0.642	**0.928**
Real + verified inpainting	18	0.735 ± 0.061	0.608 ± 0.058	0.797	0.846
Half real + top inpainting pilot	3	**0.754 ± 0.028**	**0.627 ± 0.031**	**0.814**	0.812

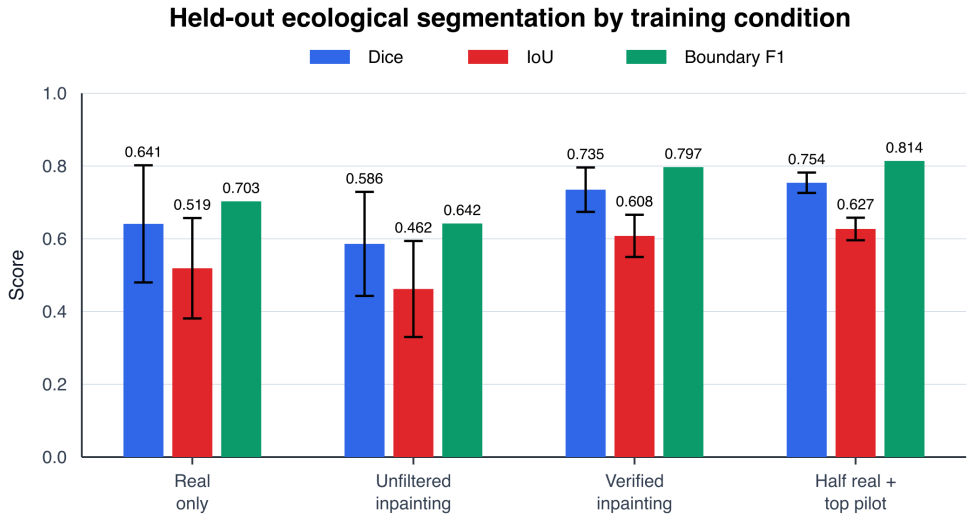

Figure 3. Held-out ecological segmentation metrics by training condition. Error bars show standard deviations where reported in the manuscript table.

6 Discussion

The synthetic examples appear to help for three reasons. First, they expose the model to ecological backgrounds that are absent from clean laboratory images. This reduces false positives on debris because the model sees more non-microplastic clutter during training. Second, they preserve pixel-level masks, so the model can learn microplastic shape while seeing new background contexts. Third, quality filtering removes failed generations before they can corrupt the training distribution. These factors together convert synthetic data from a source of noise into a controlled form of domain expansion. The improvement is greatest in boundary F1, which rises from 0.703 for real-only training to 0.797 for verified inpainting. This supports the interpretation that synthetic ecological context improves not only particle detection but also boundary

Table 2: Architecture-level Dice change after verified inpainting for the retained segmentation models.

Model	Real-only Dice	Verified-inpainting Dice	Absolute gain	Interpretation
U-Net++	0.704	0.806	+0.102	Largest gain; multiscale decoder benefits from added ecological variation
SegFormer-B2	0.728	0.789	+0.061	Transformer features improve with broader background diversity
DeepLabV3+	0.694	0.758	+0.064	Atrous context helps exploit synthetic foreground-background variation
FPN	0.685	0.735	+0.050	Moderate gain from multiscale synthetic examples
U-Net	0.660	0.710	+0.050	Improves but remains less robust than U-Net++

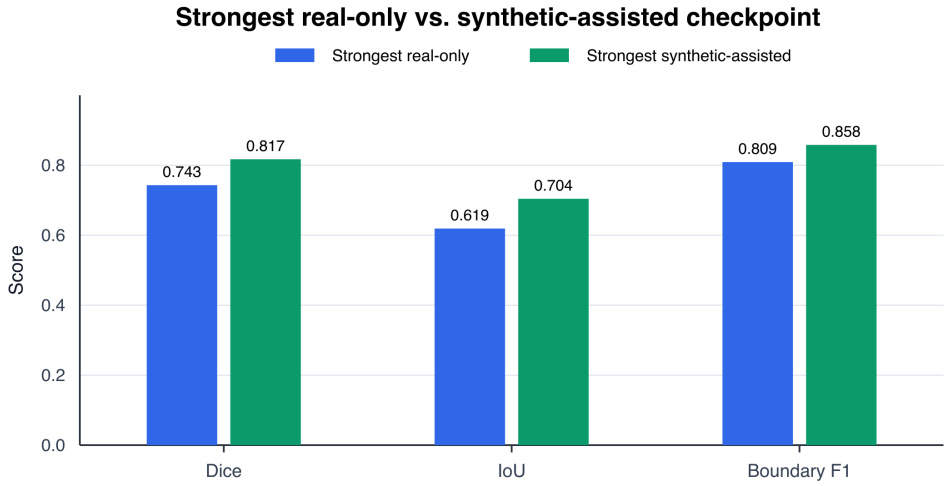

Figure 4. Strongest real-only checkpoint versus strongest synthetic-assisted checkpoint on held-out ecological evaluation metrics.

localization. That matters for microplastic analysis because area, length, and morphology are derived from the mask boundary. The half real + top inpainting pilot is consistent with this interpretation: using equal counts of real and top-ranked synthetic examples reaches Dice 0.754, exceeding both the real-only mean and the unfiltered-inpainting mean. Because this pilot is smaller than the full verified-inpainting condition, it should be interpreted as supportive rather than definitive evidence that rank-based filtering improves the training distribution.

7 Limitations

Several limitations qualify these findings. First, generated images, although visually plausible, may not capture the full variability and complexity of real-world ecological scenarios. Synthetic foregrounds may also encode generator-specific artifacts that improve benchmark performance without improving field reliability. Second, the dataset may not encompass all environmental contexts where microplastics occur, which may limit generalization to new sites, imaging devices, particle morphologies, or preparation protocols. Third, the held-out ecological test set is intentionally reserved for final evaluation, but it remains modest in size. Additional externally collected test sets would be needed to confirm robustness across laboratories, cameras, and sample preparation workflows.

8 Conclusion

This study demonstrates that quality-controlled synthetic ecological context can improve microplastic segmentation when labeled real-world data are limited. It also contributes a curated set of manually segmented microplastic microscopy images, providing paired image-mask labels for real supervised training and held-out ecological evaluation. Across the retained architectures, verified synthetic examples improved held-out ecological segmentation relative to real-only training, with the strongest synthetic-assisted model outperforming the strongest real-only model by 0.074 Dice and 0.085 IoU. The consistent gains across the retained model families support the central conclusion that synthetic ecological context can complement real labeled images when candidate generations are filtered before training. Future work should evaluate the approach across additional imaging platforms, environmental settings, and microplastic morphologies, as well as investigate more advanced generative methods and larger-scale ecological datasets. Integrating quality-controlled synthetic data with deep learning segmentation may help enable faster, more scalable, and more accessible microplastic monitoring workflows for environmental research and pollution assessment.

9 Data and Code Availability

The project code and training pipeline are available in the project GitHub repository. The associated public data resources are available through the Harvard Dataverse archive. The dataset includes the manually segmented microplastic image-mask pairs used for real supervised training and held-out ecological evaluation, along with the ecological background images used for synthetic context generation. All publication-ready results should be regenerated from the locked evaluation files and registered checkpoints.

10 Acknowledgments

We thank the Moore Institute for Plastic Pollution Research for providing microplastic imagery and supporting data resources used in this work.

11 Generative AI Statement

Generative AI tools were used to assist with preparation of this manuscript, including drafting, editing, and language refinement. The authors reviewed, revised, and approved all content and take full responsibility for the final manuscript.

References

[1] Ziani, K. et al. (2023) Microplastics: A real global threat for environment and food safety: A state of the art review. Nutrients 15(3):617. doi:10.3390/nu15030617.

[2] Stock, F. et al. (2020) Pitfalls and limitations in microplastics analyses. Plastics in the Aquatic Environment – Part I. The Handbook of Environmental Chemistry, vol. 111. Springer, Cham. doi:10.1007/698_2020_654.

[3] Primpke, S., Christiansen, S.H., Cowger, W., et al. (2020) Critical assessment of analytical methods for the harmonized and cost-efficient analysis of microplastics. Applied Spectroscopy 74(9):1012–1047. doi:10.1177/0003702820921465.

[4] Sabir, M.W. et al. (2022) Segmentation of liver tumor in CT scan using ResU-Net. Applied Sciences 12(17):8650. doi:10.3390/app12178650.

[5] Zhang, Y., Zhang, D., and Zhang, Z. (2023) A critical review on artificial intelligence-based microplastics imaging technology: Recent advances, hot-spots, and challenges. International Journal of Environmental Research and Public Health 20:1150. doi:10.3390/ijerph20021150.

[6] Geyer, R., Jambeck, J.R., and Law, K.L. (2017) Production, use, and fate of all plastics ever made. Science Advances 3:e1700782. doi:10.1126/sciadv.1700782.

[7] Hidalgo-Ruz, V., Gutow, L., Thompson, R., and Thiel, M. (2012) Microplastics in the marine environment: A review of the methods used for identification and quantification. Environmental Science & Technology 46:3060–3075. doi:10.1021/es2031505.

[8] Sherrod, H. et al. (2024) One4All: An open source portal to validate and share microplastics data and beyond. Journal of Open Source Software 9(99):6715. doi:10.21105/joss.06715.

[9] Brandt, J., Mattsson, K., and Hassellöv, M. (2021) Deep learning for reconstructing low-quality FTIR and Raman spectra: A case study in microplastic analyses. Analytical Chemistry 93:16360–16368. doi:10.1021/acs.analchem.1c02618.

[10] Yurtsever, M. and Yurtsever, U. (2019) Use of a convolutional neural network for the classification of microbeads in urban wastewater. Chemosphere 216:271–280. doi:10.1016/j.chemosphere.2018.10.084.

[11] Shorten, C. and Khoshgoftaar, T.M. (2019) A survey on image data augmentation for deep learning. Journal of Big Data 6:60. doi:10.1186/s40537-019-0197-0.

[12] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. CVPR. doi:10.1109/CVPR.2017.632.

[13] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022) High-resolution image synthesis with latent diffusion models. CVPR. doi:10.1109/CVPR52688.2022.01042.

[14] Podell, D. et al. (2023) SDXL: Improving latent diffusion models for high-resolution image synthesis. arXiv:2307.01952.

[15] Lugmayr, A. et al. (2022) RePaint: Inpainting using denoising diffusion probabilistic models. CVPR. doi:10.1109/CVPR52688.2022.01117.

[16] Suvorov, R. et al. (2022) Resolution-robust large mask inpainting with Fourier convolutions. WACV. doi:10.1109/WACV51458.2022.00323.

[17] Li, W., Lin, Z., Zhou, K., Qi, L., Wang, Y., and Jia, J. (2022) MAT: Mask-aware transformer for large hole image inpainting. CVPR. doi:10.1109/CVPR52688.2022.01049.

[18] Ronneberger, O., Fischer, P., and Brox, T. (2015) U-Net: Convolutional networks for biomedical image segmentation. MICCAI, 234–241. doi:10.1007/978-3-319-24574-4_28.

[19] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., and Liang, J. (2018) UNet++: A Nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 3–11. doi:10.1007/978-3-030-00889-5_1.

[20] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. ECCV, 833–851. doi:10.1007/978-3-030-01234-2_49.

[21] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017) Feature pyramid networks for object detection. CVPR, 2117–2125. doi:10.1109/CVPR.2017.106.

[22] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., and Luo, P. (2021) SegFormer: Simple and efficient design for semantic segmentation with transformers. NeurIPS. arXiv:2105.15203.

[23] Jocher, G. and Qiu, J. (2024) Ultralytics YOLO11. Version 11.0.0. https://github.com/ultralytics/ultralytics

[24] Yao, Y., Xu, W., and Fan, H. (2025) A Deep Learning Approach for Microplastic Segmentation in Microscopic Images. Toxics 13(12):1018. doi:10.3390/toxics13121018.

[25] Xu, J. and Wang, Z. (2024) Efficient and accurate microplastics identification and segmentation in urban waters using convolutional neural networks. Science of the Total Environment 911:168696. doi:10.1016/j.scitotenv.2023.168696.

[26] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016) V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 3DV, 565–571. doi:10.1109/3DV.2016.79.

[27] Martin, D.R., Fowlkes, C.C., and Malik, J. (2004) Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5):530–549. doi:10.1109/TPAMI.2004.1273918.

[28] Park, H.-m., Park, S., de Guzman, M.K., Baek, J.Y., Cirkovic Velickovic, T., Van Messem, A., and De Neve, W. (2022) MP-Net: Deep learning-based segmentation for fluorescence microscopy images of microplastics isolated from clams. PLOS ONE 17(6):e0269449. doi:10.1371/journal.pone.0269449.

Appendix A Model Settings

A.1 Shared Study Settings

All images were resized to 512×512 pixels. The random seeds were 13, 37, and 101. The primary evaluation split was the locked C3-clean ecological test set, and the primary threshold for binary masks was 0.5. Semantic segmentation models were trained with Dice loss plus binary cross entropy using the AdamW optimizer. The batch size was 8, the maximum training horizon was 80 epochs, the learning rate was 0.0001, and the weight decay was 0.00001. Early stopping patience was 15 epochs. AMP was enabled when CUDA was available. Synthetic examples were split into 80% training and 20% validation subsets.

A.2 Stable Diffusion Inpainting Model

The Stable Diffusion inpainting model was used for synthetic ecological microplastic image generation. The generator family was Stable Diffusion inpainting. The primary model ID was diffusers / stable-diffusion-xl-1.0-inpainting-0.1, and the Stable Diffusion 2 inpainting model ID was stabilityai / stable-diffusion-2-inpainting. The image size was 512×512 pixels.

For SDXL inpainting, the positive prompt was: "a realistic microscope ecological water sample containing one visible translucent colored microplastic fiber or fragment inside the masked region, natural debris, sharp focus." The negative prompt was: "cartoon, illustration, fake texture, text, watermark, blurry, unchanged background, empty mask, pipe, thick tube." The SDXL guidance scale was 7.0, the SDXL inference step count was 40, and the SDXL strength was 0.99. The SD2 guidance scale was 7.5, the SD2 inference step count was 45, and the SD2 strength was 0.99.

Seeded object initialization was enabled with a seeded object mix of 0.65. Diffusion blend was enabled, the diffusion mask mode was full, and inpaint masks were dilated by 4 px. Low-change generations were rejected. The minimum masked MAD was 12.0, the minimum changed-pixel fraction was 0.35, the minimum background masked MAD was 20.0, and the minimum background changed-pixel fraction was 0.50. The maximum generation attempt multiplier was 12. The insertion area fraction ranged from 0.006 to 0.05. Object opacity was 0.95, alpha feather radius was 0.7, and color jitter used brightness from 0.85 to 1.15 and contrast from 0.90 to 1.15.

A.3 Semantic Segmentation Models

A.3.1 U-Net ResNet34

The U-Net ResNet34 model was named smp_unet_resnet34. It was a semantic segmentation model implemented with segmentation-models-pytorch, using the Unet architecture with a resnet34 encoder. The model used 3 input channels and produced 1 binary foreground mask class. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.

A.3.2 U-Net++ EfficientNet-B4

The U-Net++ EfficientNet-B4 model was named smp_unetpp_effb4. It was a semantic segmentation model implemented with segmentation-models-pytorch, using the UnetPlusPlus architecture with an efficientnet-b4 encoder. The model used 3 input channels and produced 1 binary foreground mask class. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.

A.3.3 DeepLabV3+ ResNet50

The DeepLabV3+ ResNet50 model was named smp_deeplabv3plus_resnet50. It was a semantic segmentation model implemented with segmentation-models-pytorch, using the DeepLabV3Plus architecture with a resnet50 encoder. The model used 3 input channels and produced 1 binary foreground mask class. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.

A.3.4 FPN EfficientNet-B3

The FPN EfficientNet-B3 model was named smp_fpn_effb3. It was a semantic segmentation model implemented with segmentation-models-pytorch, using the FPN architecture with an efficientnet-b3 encoder. The model used 3 input channels and produced 1 binary foreground mask class. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.

A.3.5 MONAI U-Net

The MONAI U-Net model was named monai_unet. It was a semantic segmentation model implemented with MONAI, using a two-dimensional UNet architecture. The model used 3 input channels and produced 1 binary foreground mask output channel. The channel widths were 32, 64, 128, 256, and 512; the strides were 2, 2, 2, and 2; and the number of residual units was 2. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.

A.3.6 SegFormer-B2

The SegFormer-B2 model was named segformer_b2. It was a semantic segmentation model implemented with Hugging Face transformers, using the nvidia/segformer-b2-finetuned-ade-512-512 model ID. The model used 3 input channels. The pretrained ADE head was replaced with a binary segmentation head that produced 1 binary foreground mask class. The output logits were upsampled to the 512×512 mask size. Training used logits, and evaluation applied a sigmoid activation followed by the 0.5 threshold.