Automated Identification and Characterization of Particles in Holographic Images from Lake Geneva

Antoine Tissot-Favre 315999

June 6, 2025

1   Introduction

Lakes contain diverse suspended particulate matter (SPM), including biological (e.g., plankton) and mineral particles, which play critical roles in biogeochemical cycles and ecological processes [1]–[3]. Their composition influences nutrient transport, light availability, and contaminant mobility. Physical processes like flocculation can aggregate fine particles into macroflocs, enhancing settling and microbial activity [4]–[6].

Plankton, foundational to aquatic food webs, contribute significantly to global oxygen production [7] and respond rapidly to environmental changes, making them effective biosensors [8], [9]. However, in-situ analysis of such fragile and heterogeneous aggregates is challenging.

Recent advances in underwater holography, such as the LISST-Holo2 system, enable high-resolution, 3D imaging of suspended particles [10], [11]. These systems generate large volumes of data, necessitating automated methods for segmentation and classification.

In this study, we analyze holographic data from Lake Geneva to: (i) segment particles from 3D `.tiff` volumes; (ii) extract descriptive features and explore cluster structure; and (iii) evaluate supervised methods for classifying particles as Living or Non-living.

We compare morphological and deep learning-based segmentation methods, favoring classical image processing for its speed and stability. We then evaluate handcrafted and deep (DenseNet121) features for clustering (KMeans, GMM, DBScan) and classification (Random Forest, XGBoost, LightGBM), including techniques to mitigate class imbalance. Finally, we critically assess the limitations of each approach in light of the complexity and sparsity of labeled data.

2   Dataset

2.1   Sampling Site and Data Acquisition

The dataset was collected during six seasonal campaigns from the LéXPLORE platform on Lake Geneva (46°30'0.819" N, 6°39'39.007" E), a 10 m × 10 m floating station anchored 570 m offshore near Pully. Operated by a consortium of Swiss institutions, it enables high-resolution monitoring of physical, chemical, and biological processes in a 110 m water column.

Lake Geneva is a large peri-alpine lake ($580.1\,km^2$, max depth 309.7 m) fed primarily by the Rhône River, which contributes 70% of inflow and most of the suspended sediments. The river's glacial origin and hydroelectric influence produce seasonal sediment pulses with mean concentrations of 232 mg/L, dominated by clays, silts, and microflocs [5]. During stratification, the Rhône flows as an interflow at 20–40 m depth, transporting fine particles laterally. The LéXPLORE platform is thus well-positioned to study in-situ particle aggregation and flocculation [2], [4], [6].

2.2   Holographic Imaging and Dataset Structure

Suspended particles were imaged using the *LISST-Holo2* [10], a submersible digital in-line holographic camera. Each hologram consists of 101 grayscale slices (101 × 1240 × 1616 pixels, 8-bit `uint8`), with isotropic resolution of 1 μm.

Instead of processing full 3D volumes, we used 2D montage images generated by the Holobatch software (see Appendix 10), which selects the sharpest slice for each detected particle and compiles them onto a blank canvas (1240 × 1616). Metadata containing bounding box coordinates is also provided. Though Holobatch is a black-box tool, it enabled effective particle localization for downstream segmentation.

Built-in segmentation from the LISST-Holo2 was discarded due to poor mask quality and lack of interpretability (see Fig. 4). Instead, 108 holograms were acquired and a curated subset of 35 was manually segmented to support supervised tasks. The selected holograms span diverse depths and include both living (e.g., phytoplankton) and non-living (e.g., sediment) particles.

Annotations were performed using *QuPath* [12], assisted by the *Segment Anything Model (SAM)* [13], and exported in `.geojson` format for further analysis.

3 Image Analysis

3.1 Particle Segmentation

Accurate particle segmentation is a prerequisite for feature extraction and classification. Given the diversity of particle shapes and sizes, robust segmentation ensures meaningful downstream analysis.

Earlier efforts at EPFL [14] used YOLOv11 Nano for detection on montage images, combining bounding boxes with patch-wise inference. Despite reaching mDice 0.75 and recall 1.00, the method suffered from slow inference, missed detections, and coarse, square-shaped masks, as illustrated in Fig. 3. The model was trained on 35 manually annotated holograms using QuPath [12] and assisted by SAM [13].

We independently tested SAM in two modes: directly on montage images using bounding boxes as prompts, and on the original 3D slices corresponding to each particle. While the latter improved spatial context, results remained unreliable, especially for small or low-contrast particles. SAM's performance degraded in the context-poor, high-variability environment typical of in-line holography.

Due to these limitations, we shifted to a classical image processing approach focused on precision and computational efficiency.

We implemented a lightweight segmentation pipeline using morphological operations, inspired by Guo et al. [15], and built with `scikit-image`. The main steps include:

1. **Preprocessing:** Convert raw grayscale `.tiff` slices to 8-bit format.

2. **Otsu Thresholding:** Binarize using Otsu's global thresholding to separate particles from background.

3. **Morphological Closing:** Smooth contours and fill small gaps using disk-shaped structuring elements.

4. **Hole Filling:** Fill internal voids in particle regions.

5. **Anisotropic Linking:** Enhance connectivity of fragmented or elongated shapes using rotated rectangular elements at 22.5° intervals.

6. **Small Object Removal:** Eliminate objects below a size threshold to reduce noise.

7. **ROI Extraction:** For each labeled region, extract both the binary mask and the corresponding grayscale crop. The mask is applied to the crop to preserve intensity and texture within the particle's shape (Fig. 2).

The final output of the segmentation pipeline is a set of grayscale ROI images in which each particle is isolated by applying its binary mask directly to the original image. This process preserves the particle's internal texture and intensity patterns while removing background noise, resulting in clean, high-contrast representations suitable for downstream analysis. These masked grayscale ROIs serve as the input for all subsequent feature extraction, clustering, and classification steps. This approach enables efficient batch segmentation of hologram slices without requiring a GPU. Its robustness and simplicity make it particularly well-suited to the heterogeneous, low-contrast data encountered in in-line holography. A comparison of segmentation results is presented in Fig. 5. A full-scale example of the corresponding outputs applied to an entire montage image is provided in Appendix 14.
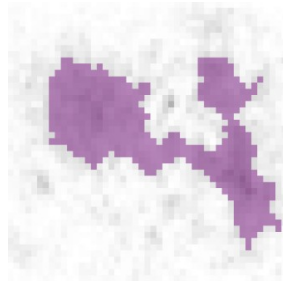
Figure 1: Original grayscale input

Figure 2: Morphological pipeline result

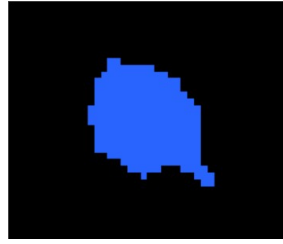Figure 3: Previous work at EPFL (YOLOv11 Nano + SAM)

Figure 4: Holobatch segmentation (built-in)

Figure 5: Comparison of segmentation results. (a) shows the original grayscale particle slice. (b) is the final result from our morphological segmentation pipeline. (c) shows the segmentation obtained using the previous deep learning approach at EPFL. (d) displays the output of the Holobatch built-in tool. The morphological method preserves contours and texture faithfully without requiring GPU inference.

3.2 Feature Extraction

To support unsupervised clustering, we extracted a diverse set of handcrafted features from each segmented ROI, consisting of a binary mask and its corresponding grayscale crop. These features captured morphology, contour geometry, texture, and intensity.

Shape and Geometric Features. Using `regionprops` and `regionprops_table` from `scikit-image`, we computed standard descriptors including area, perimeter, eccentricity, solidity, extent, orientation, and axis lengths. Convexity was calculated as the ratio of convex area to actual area.

Compactness Metrics. To assess contour complexity:

- **Alpha Shape Compactness (C_a):** Computed from an α-shape polygon built via Delaunay triangulation: $C_a = 4\pi A_a/P_a^2$.

- **Convex Hull Compactness (C_h):** Calculated similarly using the convex hull: $C_h = 4\pi A_h/P_h^2$.

- The hull-to-alpha area ratio (A_h/A_a) was also included to quantify boundary irregularity.

Fourier Descriptors. To encode global shape, we applied a Discrete Fourier Transform to each particle contour. We retained the first $n = 3$ amplitude and phase components (after removing

the DC term) to capture coarse shape while avoiding instability. Higher-order components often produced unreliable or NaN values, especially for low-resolution masks.

Hu Moments. Seven log-transformed Hu moments were computed from normalized central moments. These are invariant to scale, rotation, and translation, and commonly used for shape classification.

Texture and Intensity Features. We computed Local Binary Patterns (LBP) using a uniform 8-neighbor circular kernel (radius 1), and extracted their mean and standard deviation. Additionally, we calculated five intensity-based statistics from each grayscale ROI: mean, standard deviation, skewness, kurtosis, and Shannon entropy.

Final Feature Set. Each particle was described by:

- **Geometric:** area, perimeter, eccentricity, solidity, extent, orientation, axis lengths, convexity

- **Compactness:** C_a, C_h, A_h/A_a

- **Fourier:** amplitudes and phases of first 3 descriptors

- **Hu Moments:** 7 log-transformed values

- **LBP:** mean and standard deviation

- **Intensity:** mean, std, skewness, kurtosis, entropy

This feature set provided a rich and interpretable representation of particle shape and appearance, forming the basis for clustering and classification experiments.

3.3 Clustering

This analysis aimed to assess whether particles could self-organize into meaningful categories (e.g., biological vs. non-biological) based solely on the extracted features. Despite a systematic pipeline involving preprocessing, dimensionality reduction, filtering, and multiple clustering strategies, results consistently failed to yield interpretable or biologically meaningful groupings.

Methodological Summary
Feature Composition. The final feature set integrated morphological descriptors, intensity-based statistics, shape metrics (Hu moments, convexity, Fourier descriptors), and texture features (LBP).
Normalization. All numeric features were standardized using `StandardScaler`, except for Fourier phase components, which were preserved in raw form to maintain angular periodicity.
Dimensionality Reduction. Principal Component Analysis (PCA) was applied to reduce the feature space while preserving most of its variance. We retained enough components to explain at least 90% of the total variance, ensuring that the majority of information was maintained while mitigating redundancy.
Filtering. To remove poorly defined or extreme particles, we applied an area filter retaining only particles with areas between 80 and 2000 pixels. These thresholds were determined based on visual inspection of the segmented ROIs and analysis of the area distribution histogram (Appendix Fig. 17), aiming to exclude both under-segmented noise and large outliers likely due to merging or artifacts.

Clustering Performance To assess the presence of structure in the extracted feature space, we applied three clustering algorithms: **KMeans** (distance-based), **Gaussian Mixture Models (GMM)** (probabilistic soft clustering), and **DBSCAN** (density-based). Each method was tested on both the unfiltered dataset and a filtered version where small, low-information particles were removed.

Table 1: Clustering metrics across algorithms and datasets. Silhouette and CH scores reflect cluster compactness and separation; lower DBI indicates better clustering quality.

Method	Dataset	Silhouette	Davies-Bouldin ↓	Calinski-Harabasz ↑
KMeans	Unfiltered	0.210	1.544	4979.6
	Filtered	0.204	1.550	991.1
GMM	Unfiltered	0.100	2.004	3100.5
	Filtered	0.060	2.512	324.6
DBSCAN	Unfiltered	-0.440	1.381	7.5
	Filtered	–	–	–

As shown in Table 1, all three clustering models exhibit poor performance. Silhouette scores remain below 0.25 across the board, indicating weak and overlapping clusters. The Davies-Bouldin Index (DBI)[1] exceeds 1.5 in all cases, while the Calinski-Harabasz Index (CH)[2] drops significantly in the filtered dataset. GMM, despite its flexibility, performed worse than KMeans in all metrics. DBSCAN failed to identify any clusters in the filtered data and assigned over 90% of points in the unfiltered set as noise. These results confirm that the feature space lacks meaningful separable structure and that unsupervised clustering is not viable under the current representation.

Qualitative Evaluation UMAP Projections (see Fig. 6) revealed diffuse, overlapping clusters with no clear separation, suggesting the absence of discrete grouping.
Silhouette Histograms (Fig. 7) confirmed these findings, with most values concentrated near zero and many negative scores, indicating misclassification or overlapping assignments.
ROI Cluster Visualizations (Figs. 16) display randomly sampled Regions of Interest from each cluster, for both the unfiltered and filtered datasets. These grids were intended to assess intra-cluster visual consistency and reveal potential shared biological or morphological traits.

- **High Intra-Cluster Variability:** Within most clusters, particles exhibit a wide range of shapes, sizes, and grayscale patterns. This is particularly pronounced in the unfiltered case, where poorly segmented or low-contrast ROIs are common.

- **Weak Visual Coherence:** While a few clusters show slight trends (e.g., Cluster 0 in the unfiltered case contains several elongated structures), no cluster forms a clearly homogeneous group. The presence of circular, elongated, and amorphous structures within the same cluster suggests that the grouping is driven by subtle differences in a few features rather than consistent morphology.

- **Filtering Improves Clarity but Not Consistency:** The filtered dataset, having removed extremely small and large ROIs, yields visually cleaner examples, but intra-cluster coherence remains low. Random samples still include both complex planktonic-like structures and blurry or featureless blobs within the same cluster.

---

[1]Lower DBI values indicate more compact, well-separated clusters.
[2]Higher CH values reflect stronger inter-cluster separation.

- **Ambiguity Across Clusters:** Some ROIs (e.g., small dense spots or fuzzy diffuse regions) appear across multiple clusters, reinforcing the hypothesis that the feature space lacks strong discriminative power.

In conclusion, ROI visualizations confirm the clustering metrics: clusters are not meaningfully distinct. Even with random sampling, each cluster contains a wide variety of shapes and textures, reinforcing that unsupervised clustering of handcrafted features does not yield biologically relevant groupings.

Radar plots (Appendix Fig. 15) show that clusters differ mainly in brightness-related features (e.g., intensity mean, standard deviation), with some contribution from shape descriptors like Hu moments and hull perimeter. However, these differences are too weak to produce coherent or interpretable clusters, as further evidenced by both metrics and visual inspection.
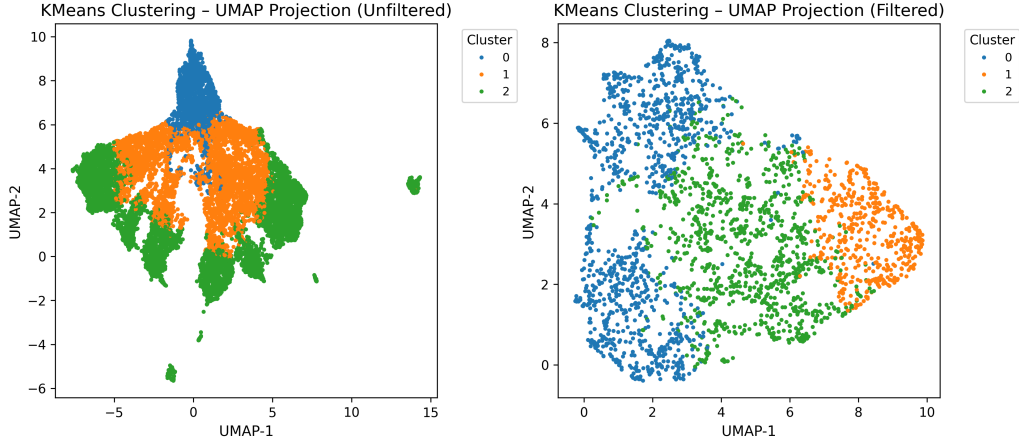

Figure 6: UMAP projection of particle features clustered using KMeans (filtered and unfiltered datasets). Clusters are highly overlapping with no clear separation.

(a) Unfiltered dataset
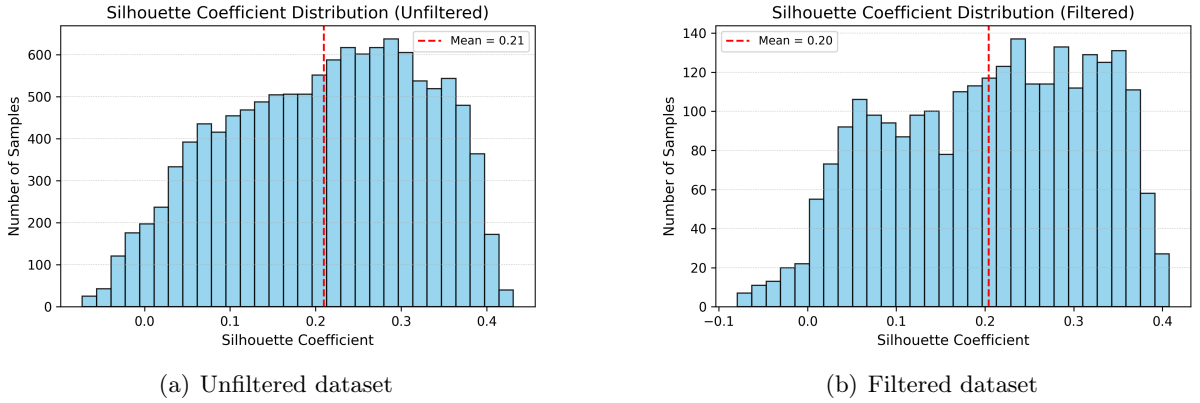
(b) Filtered dataset

Figure 7: Silhouette coefficient distributions for KMeans clustering. Most samples fall near zero or below, indicating poor cluster definition.

Why Clustering Fails   Several factors explain the poor performance of clustering. First, the extracted features, such as Hu moments, convexity, and LBP, only partially align with biological or physical differences between particle types like phytoplankton, zooplankton, and sediment. Although they capture some structural variation, this is not consistent enough to drive clear separation. Second, in-line holography introduces imaging artifacts such as fringe

6

patterns, low contrast, and partial reconstructions, which degrade signal quality and reduce feature reliability. Additionally, the boundary between living organisms and non-living particles appears continuous in feature space, resulting in overlapping groups with no sharp distinction. Finally, the absence of labeled ground truth prevents validation or refinement of clusters, further limiting their interpretability.

Final Assessment   In summary, despite extensive preprocessing and feature engineering, the dataset does not support meaningful unsupervised clustering. While simple traits like area and brightness drive some variability, they are insufficient to form coherent or biologically interpretable groups. This is reflected in low silhouette scores, high Davies-Bouldin indices, and the lack of consistent structure in visualizations such as UMAP and silhouette plots. High intra-cluster variability and weak separation reinforce this conclusion. Future work should therefore focus on richer representations, such as those derived from deep or self-supervised models; better segmentation to improve feature consistency; and expert-labeled datasets to enable supervised classification. Based on these insights, we transition in the next section to a supervised learning framework to assess the discriminative power of known labels.

3.4   Living / Non-Living Classification

Motivation and Transition   Due to the lack of meaningful clusters found through unsupervised approaches, we shifted to supervised learning. While the dataset lacks detailed semantic labels (e.g., distinguishing plankton from sediment), it does include binary annotations, *Living* vs. *Non-Living*, provided by limnology experts.
We aimed to determine whether the same handcrafted features that failed in clustering could nonetheless support classification between these two broad categories. A positive result would suggest that supervised models can capture subtle, high-dimensional patterns not easily revealed through unsupervised techniques.

Labeled Dataset and Class Imbalance   The labeled dataset included 404 particles: 81 labeled as *Living* and 323 as *Non-Living*, resulting in a strong class imbalance (20% minority class). This imbalance, common in freshwater particulate datasets, poses challenges for classification models, which often default to the majority class without appropriate correction strategies.

Feature Selection and Preprocessing   We used the same handcrafted features described in Section 3.2, along with the same preprocessing pipeline. Features were standardized using `StandardScaler`, and Fourier phase components were retained in raw form due to their angular nature. No additional feature engineering or selection was performed at this stage to maintain comparability with the clustering setup.
To address class imbalance, we explored two standard balancing techniques:

- **SMOTE (Synthetic Minority Oversampling Technique):** A synthetic oversampling method that generates new examples of the minority class by interpolating between existing samples. SMOTE has been widely used for balancing binary datasets and is known to reduce overfitting compared to naive duplication [16].

- **Undersampling:** We randomly downsampled the majority class (Non-Living) to match the number of minority class samples (Living), yielding a balanced training set. This approach sacrifices data volume in exchange for balance and simplicity.

Both techniques were applied after stratified splitting of the dataset into training and test subsets, using an 80/20 split.
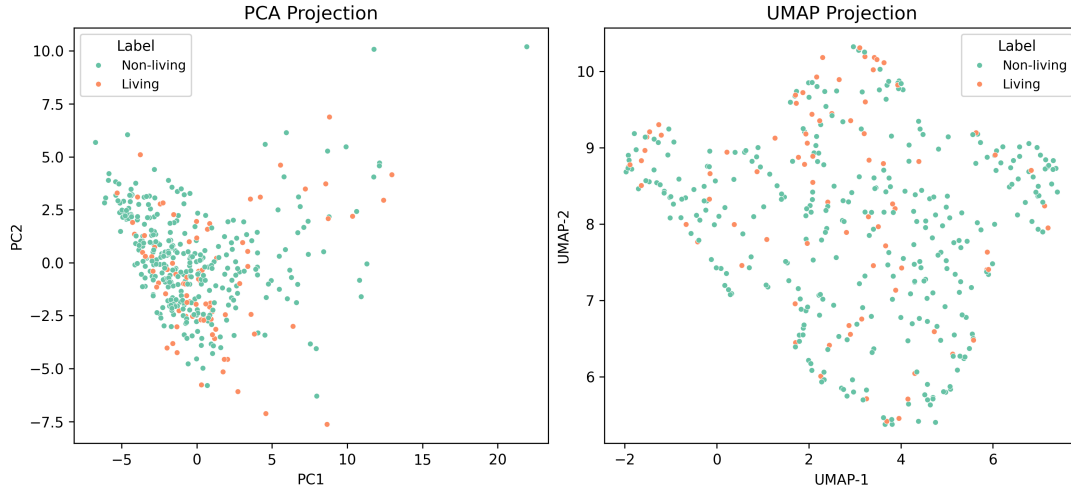
Figure 8: Projection of labeled data using PCA (left) and UMAP (right), color-coded by class. The Living and Non-Living classes show substantial overlap in both low-dimensional embeddings.

Model Selection and Justification   Given the observed class overlap in feature space (Fig. 8), we selected models that are robust to feature interactions, class imbalance, and noise. Specifically, we chose:

- **Random Forest:** A bagging ensemble of decision trees that reduces variance and provides feature importance estimates. It handles non-linear interactions and is robust to overfitting on small datasets.

- **XGBoost:** A gradient-boosted tree method optimized for speed and regularization [17]. XGBoost is known for its high performance on tabular data, and it supports class weighting natively.

- **LightGBM:** A gradient boosting framework that uses histogram-based learning for efficiency [18]. It is particularly well-suited to large feature spaces and can also handle imbalance through built-in class weighting.

These models were selected not only for their robustness to noisy or imbalanced data, but also for their ability to reveal which features contribute most to classification, an important aspect for scientific interpretability.

Evaluation and Discussion   Classifying Living versus Non-Living particles based on hand-crafted features proved difficult. Across all tested models: Random Forest, XGBoost, and LightGBM, performance varied significantly with the resampling strategy (none, SMOTE, or undersampling), but no configuration achieved robust separation between classes.

LightGBM without resampling unexpectedly achieved the best performance (84% accuracy and 0.43 F1-score for the minority class). However, training logs frequently issued the warning *"No further splits with positive gain"*, suggesting the model struggled to find informative feature splits and likely defaulted to weak but consistent rules. XGBoost with undersampling achieved the highest recall (56%) for the minority class, though at the cost of reduced accuracy (57%). This trade-off may be acceptable in scenarios prioritizing sensitivity to Living particles. Random Forest performed worst overall, especially without balancing, though recall improved to 44% under undersampling.

These trends, summarized in Table 2, highlight the main limitation: the feature space does not enable stable or reliable discrimination. PCA and UMAP projections confirm substantial

overlap between classes. Intraclass variability among living particles likely exceeds interclass differences, and the handcrafted descriptors fail to capture the biological patterns needed for clean separation. Resampling strategies offered mixed benefits: undersampling boosted minority recall but degraded overall performance, while SMOTE introduced instability, sometimes degrading minority precision and F1 scores.

Table 2: Classification results across models and resampling strategies. Metrics shown for the minority class (Living). *Note: the test set included only 16 Living samples, making scores sensitive to prediction changes.*

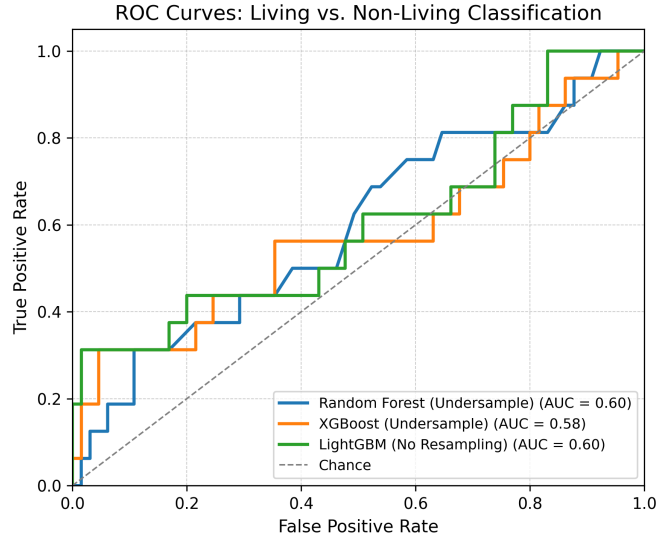| Model	Balancing	Acc.	F1 (Living)	Prec. (Living)	Rec. (Living)
Random Forest	None	0.81	0.21	0.67	0.12
	SMOTE	0.73	0.21	0.25	0.19
	Undersample	0.63	0.32	0.25	0.44
XGBoost	None	0.81	0.40	0.56	0.31
	SMOTE	0.75	0.33	0.36	0.31
	Undersample	0.57	0.34	0.24	0.56
LightGBM	None	**0.84**	**0.43**	**0.71**	**0.31**

Figure 9: ROC curves for the best configurations of each model: Random Forest (undersampling), XGBoost (undersampling), and LightGBM (no resampling). All curves remain close to the diagonal, indicating limited separability of Living vs. Non-Living particles.

In conclusion, none of the models achieved consistent or interpretable performance. The low F1-scores for the minority class (0.21–0.43), combined with unstable learning dynamics and weak ROC behavior, suggest that handcrafted features lack the resolution needed to support binary classification in this context. Improvements will likely require better representations (e.g., through deep learning), expanded annotation efforts, and potentially a richer class hierarchy beyond the binary Living/Non-Living split.

3.5 Learned Feature Representations

Motivation The previous sections demonstrated that handcrafted features, while capturing basic morphological, textural, and intensity-related aspects of the particles, were insufficient to enable robust clustering or classification. Models struggled to generalize across the Living vs.

Non-Living boundary, and PCA/UMAP projections confirmed substantial overlap between the two classes.

These limitations highlight the need for richer, more abstract representations. Deep neural networks, especially convolutional architectures, have shown strong capabilities in extracting high-level visual features from images, even in noisy or under-labeled domains. In particular, pre-trained models offer a way to leverage prior visual knowledge without requiring large amounts of labeled data. Thus, we explore whether learned features extracted from deep models can better capture the visual structure of the particles.

Deep Features from DenseNet121 *Extraction Pipeline.* To extract deep representations, we used DenseNet121 [19], a convolutional neural network known for efficient feature reuse and compact representation. We employed the version available in `torchvision`, removing the classification head and using the penultimate layer as a feature extractor. Each grayscale ROI was replicated across three channels, zero-padded to 224×224 pixels centered on the object, and normalized using ImageNet statistics. The resulting tensor was passed through the model in evaluation mode, and the penultimate output was flattened and stored as the learned feature vector. The feature matrix was subsequently standardized using `StandardScaler`.

Feature Space Structure. Visualizations of the learned features using PCA and UMAP (Appendix Fig. 18(a)) revealed significant overlap between classes. Although some compression was observed compared to handcrafted features, Living and Non-Living particles still did not form separable clusters. Quantitative distance analysis (Table 3) supported this: intra-class distance for Living particles (48.78) exceeded their inter-class distance (43.45), suggesting high internal variability. Non-Living particles were more compact (35.03 intra-class distance).

Table 3: Intra-class and inter-class distances in DenseNet121 feature space.

Metric	Distance
Intra-class (Non-Living)	35.03
Intra-class (Living)	48.78
Inter-class	43.45

Classification Results. Classifiers were trained on the extracted features using different sampling strategies (Table 4). Most models exhibited strong bias toward the Non-Living class. The highest recall for Living particles (56%) was achieved by Random Forest with undersampling, though with significant accuracy loss. LightGBM often failed to find meaningful splits, indicated by warnings of "no further splits with positive gain."

Table 4: Classification performance using DenseNet121 features.

Model	Sampling	Accuracy	F1 (Living)	Recall (Living)	Precision (Living)
Random Forest	None	0.80	0.00	0.00	0.00
	SMOTE	0.68	0.00	0.00	0.00
	Undersample	0.64	0.38	0.56	0.29
XGBoost	SMOTE	0.68	0.12	0.11	0.14
	Undersample	0.57	0.24	0.33	0.19
LightGBM	None	0.73	0.14	0.11	0.20
	SMOTE	0.70	0.32	0.33	0.30
	Undersample	0.80	0.00	0.00	0.00

Takeaway. DenseNet121 representations offer slightly better compression of particle variation but fail to uncover biologically meaningful separability. The Living class remains especially diverse,

making boundary learning difficult under current labels.

Self-Supervised Features with DINO *Extraction.* To test the benefit of unsupervised domain-agnostic learning, we applied DINOv2 ViT-S/14 [20], a self-supervised vision transformer trained on natural images. ROI crops were passed through the frozen DINO encoder, and the output embeddings were extracted as feature vectors.

Feature Space Structure. UMAP projections (Appendix Fig. 18(b)) showed slightly more compact clusters than with DenseNet121, but no clear Living vs. Non-Living separation emerged. Distance metrics (Table 5) were similar in magnitude to those from DenseNet, with inter-class distance only marginally exceeding intra-class values.

Table 5: Intra-class and inter-class distances in the DINO feature space.

Metric	Distance
Intra-class (Non-Living)	23.76
Intra-class (Living)	28.78
Inter-class	27.43

Classification Results. Table 6 summarizes classifier performance. Random Forest with SMOTE achieved the best F1 and recall for Living particles, but overall scores remained modest. No model exceeded 30% F1 for the minority class.

Table 6: Classification performance using DINO features.

Model	Sampling	Accuracy	F1 (Living)	Recall (Living)
Random Forest	None	0.80	0.00	0.00
	SMOTE	0.77	0.29	0.22
	Undersample	0.57	0.17	0.22
XGBoost	None	0.70	0.13	0.11
	SMOTE	0.68	0.22	0.22
	Undersample	0.50	0.15	0.22
LightGBM	None	0.75	0.15	0.11
	SMOTE	0.68	0.12	0.11
	Undersample	0.80	0.00	0.00

Takeaway. While DINO embeddings slightly improve Living class recall, they do not achieve meaningful class separation. The results suggest that self-supervised representations remain limited in capturing domain-specific biological variation when used out-of-the-box.

Comparative Summary & Final Insights DenseNet121 and DINO offered only marginal improvements over handcrafted features, with DINO showing slightly better recall and compactness. However, neither model succeeded in producing separable feature spaces or reliable classification, largely due to the high intra-class variability among Living particles and significant overlap between classes. These limitations highlight that the issue lies less in model architecture and more in the semantic ambiguity of the current labeling scheme.

Overall, pre-trained visual features, whether supervised or self-supervised, are insufficient without domain adaptation. Future work should focus on fine-tuning models on plankton-specific data, applying self-supervised methods directly on holographic images, and refining labels to better reflect biological diversity. Expanding the dataset and incorporating additional modalities such as phase or temporal information may also help resolve structural ambiguities and improve classification performance.

4 Conclusion

This study presented a comprehensive analysis pipeline for segmenting, describing, and classifying suspended particles imaged by an in-line holographic camera in Lake Geneva. Beginning with robust morphological segmentation, we extracted a rich set of handcrafted features to explore unsupervised clustering and supervised classification of particles into Living and Non-Living categories.

Unsupervised methods, including KMeans, GMM, and DBSCAN, failed to uncover meaningful structure in the feature space, with clustering metrics and visual inspection confirming the absence of coherent groupings. Supervised classifiers (Random Forest, XGBoost, LightGBM) trained on the same handcrafted features achieved moderate accuracy but struggled with poor recall and F1-scores for the minority class, revealing high intra-class variability and significant overlap between categories.

To address these limitations, we explored learned representations from DenseNet121 and the self-supervised DINO model. While these deep features compressed the feature space more effectively than handcrafted ones, they did not resolve the core challenge: the absence of clear semantic boundaries between particle types. Even with class balancing techniques, classification remained inconsistent, with no model delivering reliable separation or strong recall for Living particles.

Taken together, these findings highlight three key obstacles: (1) the intrinsic ambiguity of the Living vs. Non-Living dichotomy, (2) the limited alignment between generic visual features and biological particle semantics, and (3) the scarcity and imbalance of labeled data.

Future Work. To move beyond these limitations, we recommend the following directions:

- **Domain adaptation:** Fine-tuning deep networks on planktonic datasets or using domain-specific architectures may yield more relevant feature spaces.

- **Self-supervised pretraining:** Leveraging large volumes of unlabeled holographic data with techniques like DINO or SimCLR could better capture latent biological structure.

- **Data expansion:** More annotated examples, especially from rare or ambiguous categories—are needed to support robust learning.

- **Multimodal analysis:** Incorporating temporal, spectral, or 3D spatial context may enhance discrimination in otherwise overlapping visual data.

In summary, while automated holographic particle analysis remains promising, this work illustrates that careful integration of segmentation, representation, and domain knowledge is essential to realize its full potential in ecological monitoring and beyond.

5 Code Availability

All code and resources used in this project are available on GitLab: `https://gitlab.epfl.ch/center-for-imaging/enac/LIMNC/lake_snow_semester_project`

6 AI Disclosure

This report benefited from the use of large language models (LLMs) during its preparation. Specifically, GPT-4o (OpenAI) was used to:

- Suggest and iterate on the overall structure of the report;

- Improve technical writing clarity and conciseness;

- Format LaTeX and refine section transitions;

- Support code structure design and debugging in Python.

All scientific analyses, code implementations, and result interpretations were conducted independently.
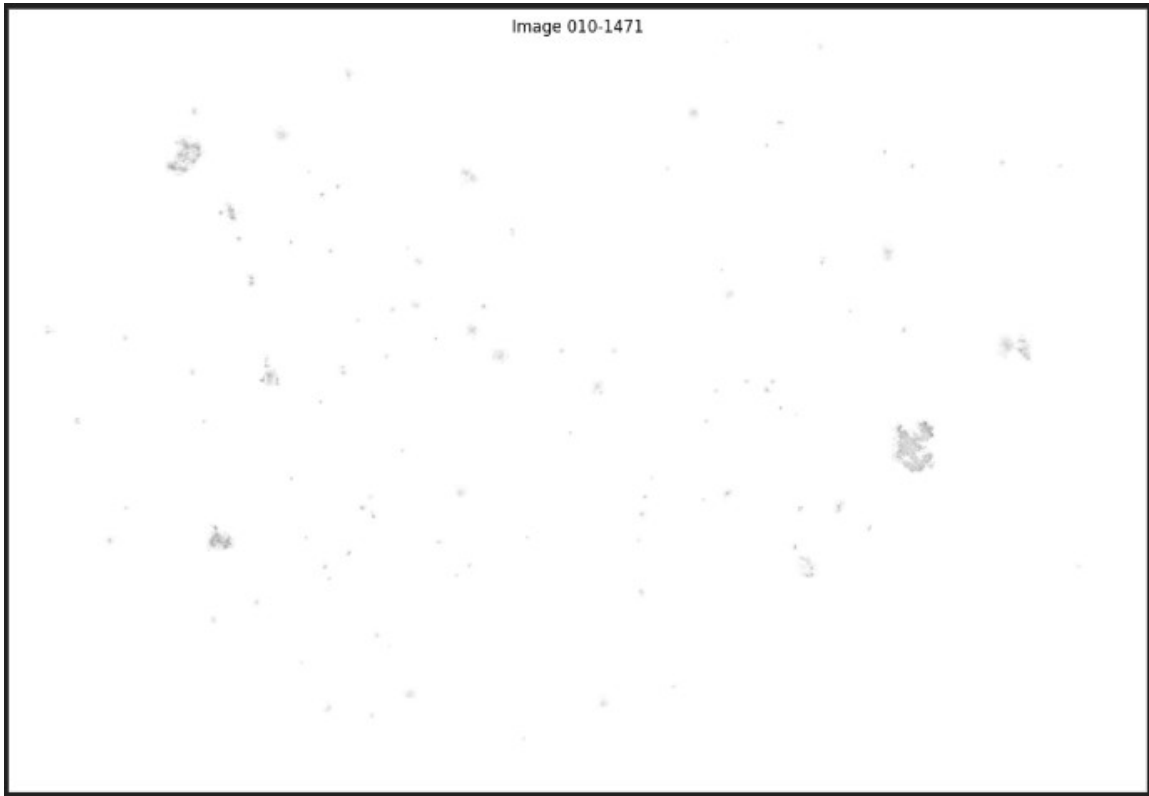
A   Supplementary Figures

Figure 10: Example of a full montage image generated by Holobatch, used as the 2D input format for all downstream segmentation tasks. Each particle is placed on a blank canvas based on its sharpest slice from the original 3D hologram.
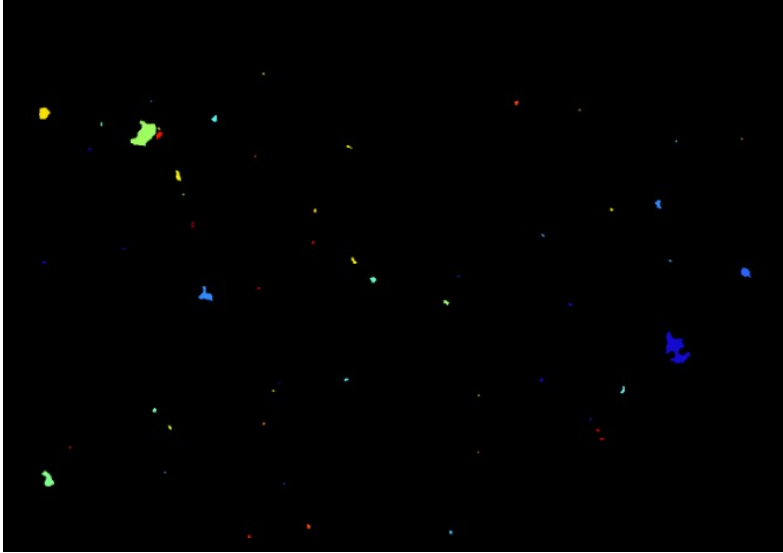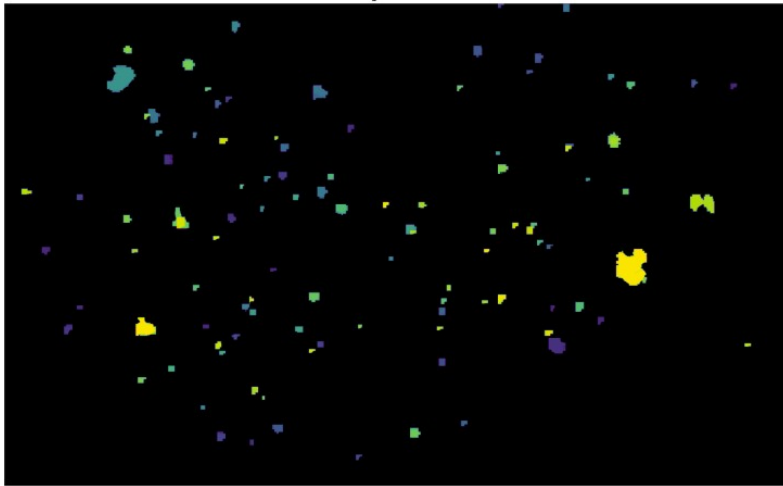
Figure 11: Holobatch output

Figure 12: Deep Learning approach
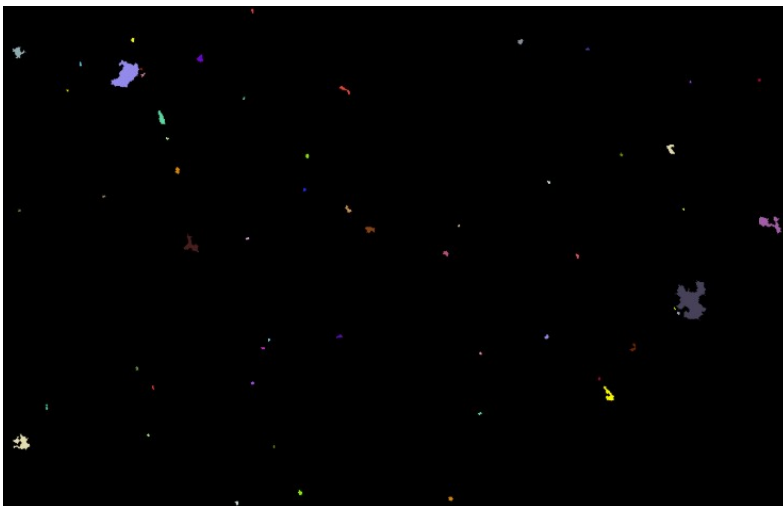
Figure 13: Morphological pipeline

Figure 14: Segmentation results on a full montage image. (a) Built-in segmentation from Holobatch. (b) Deep learning pipeline used in previous work at EPFL. (c) Result from our morphological pipeline.

(a) Unfiltered dataset
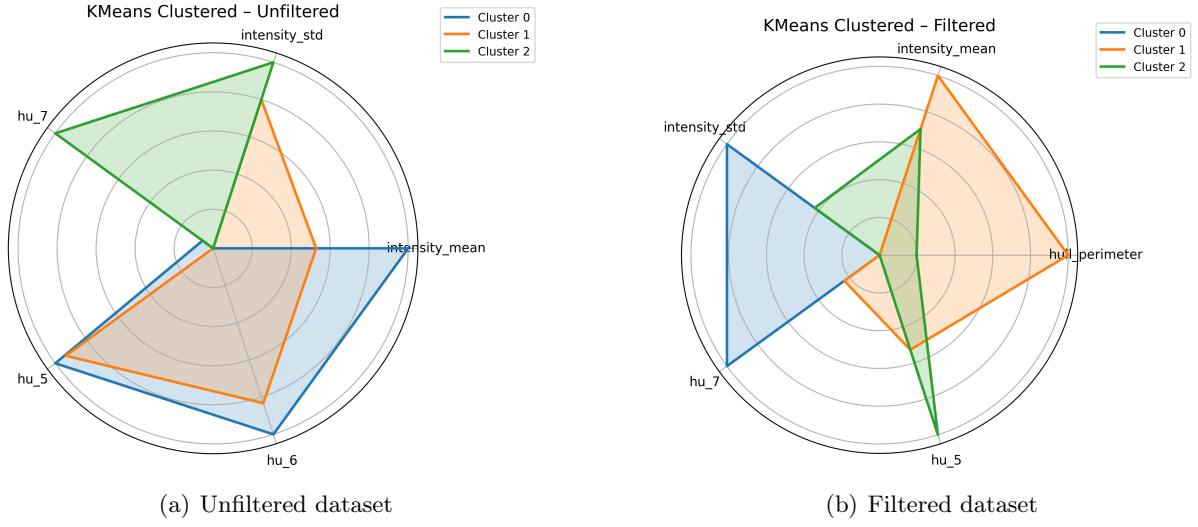
(b) Filtered dataset

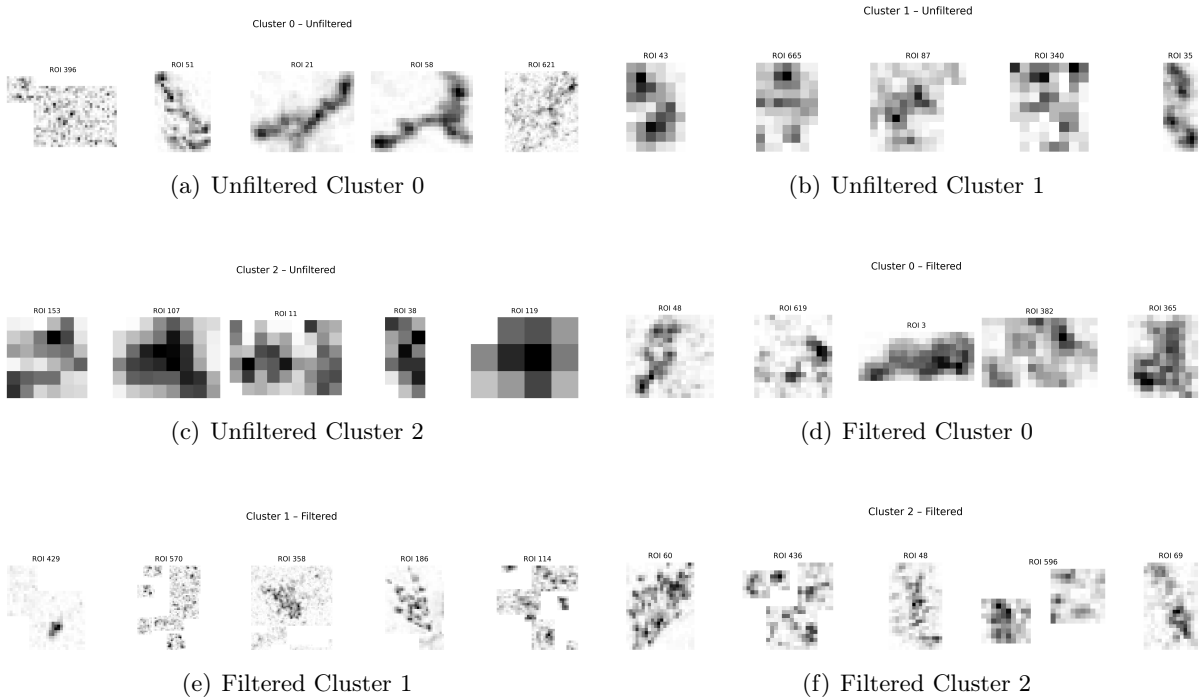Figure 15: Radar plots comparing the top-variance features across clusters for KMeans on the filtered and unfiltered datasets.

(a) Unfiltered Cluster 0

(b) Unfiltered Cluster 1

(c) Unfiltered Cluster 2

(d) Filtered Cluster 0

(e) Filtered Cluster 1

(f) Filtered Cluster 2

Figure 16: Representative ROIs from each KMeans cluster for both unfiltered and filtered datasets. Each panel shows a grid of randomly selected ROIs belonging to one cluster. The top row shows unfiltered clusters (0–2), and the bottom row shows filtered clusters (0–2).
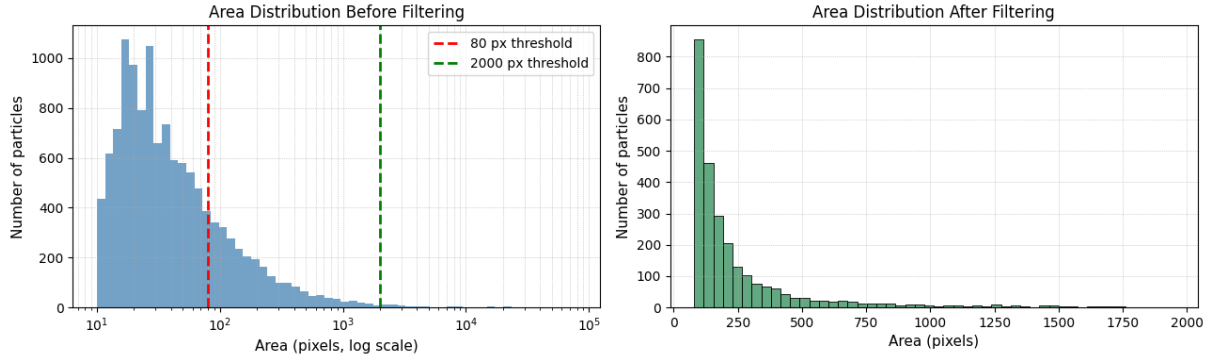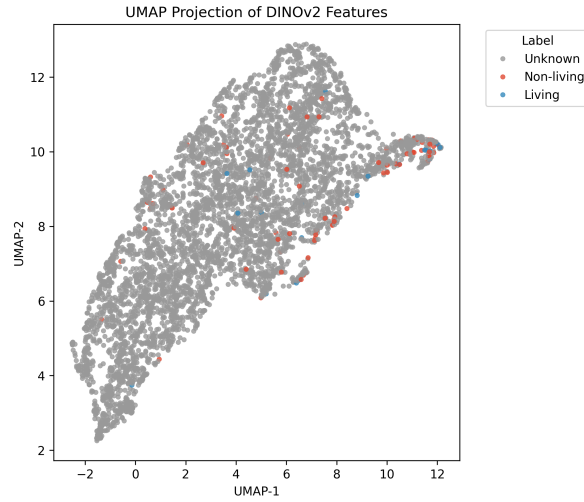
Figure 17: Distribution of particle areas before and after filtering. The applied thresholds at 80 and 2000 pixels were selected to exclude small noisy detections and large merged artifacts.
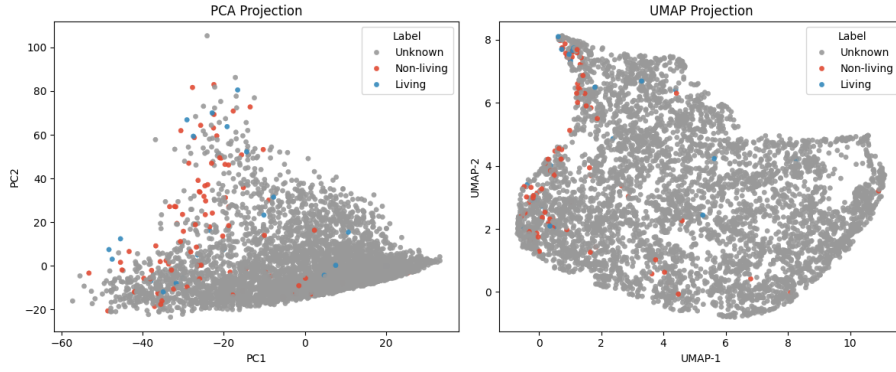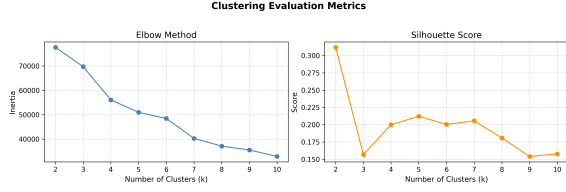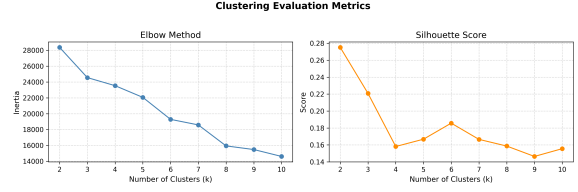

(a) DenseNet121 features

(b) DINO features

Figure 18: PCA and UMAP projections of DenseNet121 and DINO features, color-coded by Living/Non-Living labels. Both representations show high overlap, suggesting weak class separability.
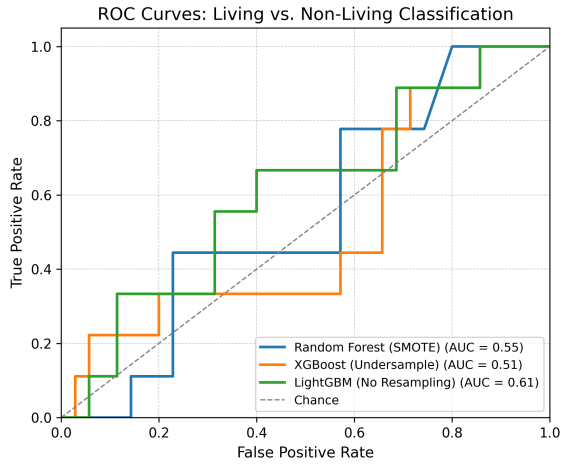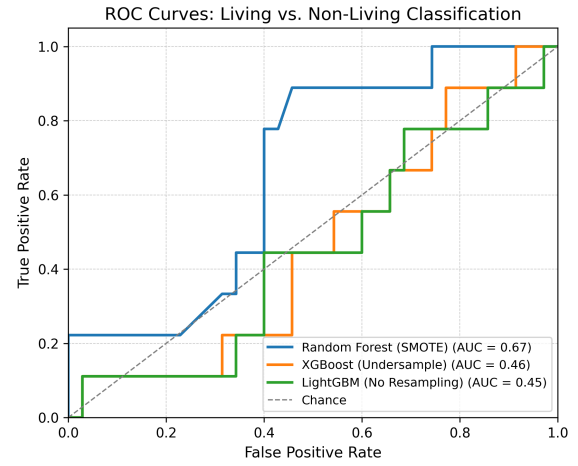
(a) DenseNet121 features

(b) DINO features

Figure 19: Silhouette scores and elbow method plots for DenseNet121 and DINO features. No strong cluster structure is apparent in either case.

(a) DenseNet121 features

(b) DINO features

Figure 20: ROC curves for classifiers trained on DenseNet121 and DINO features. All curves remain close to the diagonal, indicating limited separability between classes.

References

[1] J. M. Malmaeus and L. Håkanson, "A dynamic model to predict suspended particulate matter in lakes," *Ecological Modelling*, vol. 167, no. 3, pp. 247–262, Sep. 2003, ISSN: 0304-3800. DOI: `10.1016/S0304-3800(03)00166-2`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0304380003001662`.

[2] H.-P. Grossart and M. Simon, "Limnetic macroscopic organic aggregates (lake snow): Occurrence, characteristics, and microbial dynamics in lake constance," *Limnology and Oceanography*, vol. 38, no. 3, pp. 532–546, 1993, ISSN: 1939-5590. DOI: `10.4319/lo.1993.38.3.0532`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.4319/lo.1993.38.3.0532`.

[3] M. Simon, H.-P. Grossart, B. Schweitzer, and H.-P. Ploug, "Microbial ecology of organic aggregates in aquatic ecosystems," *Aquatic Microbial Ecology*, vol. 28, pp. 175–211, 2002, ISSN: 0948-3055. DOI: `10.3354/ame028175`. [Online]. Available: `http://www.int-res.com/abstracts/ame/v28/n2/p175-211/`.

[4] Z. Deng, Y. Yang, X. Wang, T. Li, and W. Fu, "Review of the action of organic matter on mineral sediment flocculation," *Frontiers in Earth Science*, vol. 10, 2022. DOI: `10.3389/feart.2022.965919`. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/feart.2022.965919/full`.

[5] V. Piton, J.-L. Loizeau, S. Delandmeter, *et al.*, "From particles to flocs: Revealing where flocculation occurs in the nearfield of a negatively-buoyant river plume in a large lake (lake geneva)," *Journal of Geophysical Research: Oceans*, vol. 129, no. 2, 2024. DOI: `10.1029/2023JC019860`. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1029/2023JC019860`.

[6] A. Gallorini and J.-L. Loizeau, "Lake snow as a mercury methylation micro-environment in the oxic water column of a deep peri-alpine lake," *Chemosphere*, vol. 299, p. 134 306, 2022, ISSN: 0045-6535. DOI: `10.1016/j.chemosphere.2022.134306`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0045653522007998`.

[7] H. M. Sosik and R. J. Olson, "Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry," *Limnology and Oceanography: Methods*, vol. 5, no. 6, pp. 204–216, 2007.

[8] P. F. Culverhouse, R. Ellis, R. Simpson, R. Williams, R. Pierce, and J. Turner, "Automatic categorisation of five species of cymatocylis (protozoa, tintinnida) by artificial neural network," *Marine Ecology Progress Series*, pp. 273–280, 1994.

[9] Q. Hu and C. Davis, "Automatic plankton image recognition with co-occurrence matrices and support vector machine," *Marine Ecology Progress Series*, vol. 295, pp. 21–31, 2005.

[10] Sequoia Scientific Inc., *Lisst holo2 user's manual version 1.0*, 2018. [Online]. Available: `https://www.sequoiasci.com/wp-content/uploads/2013/07/LISST-Holo2-Users-Manual-Version1.0.pdf`.

[11] A. R. Nayak, J. M. Graff, C. S. Yentsch, J. C. Kennedy, and S. L. Smith, "A review of holography in the aquatic sciences: In situ characterization of particles, plankton, and small scale biophysical interactions," *Frontiers in Marine Science*, vol. 7, 2021. DOI: `10.3389/fmars.2020.572147`. [Online]. Available: `https://www.frontiersin.org/articles/10.3389/fmars.2020.572147/full`.

[12] P. Bankhead, M. B. Loughrey, J. A. Fernández, *et al.*, "Qupath: Open source software for digital pathology image analysis," *Scientific Reports*, vol. 7, no. 1, pp. 1–7, 2017.

[13] A. Kirillov, E. Mintun, N. Ravi, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[14] V. Aolaritei, A. De Laurentis, and M. L. Le Bras, *Plankton detection and segmentation in lake geneva using holographic imaging with yolo11 nano model*, Machine Learning Course Project, EPFL, 2023. [Online]. Available: `https://github.com/CS-433/ml-project-2-mva`.

[15] B. Guo, L. Nyman, A. R. Nayak, *et al.*, "Automated plankton classification from holographic imagery with deep convolutional neural networks," *Limnology and Oceanography: Methods*, vol. 19, pp. 21–36, 2020. DOI: `10.1002/lom3.10402`.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[17] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, ACM, 2016, pp. 785–794. DOI: `10.1145/2939672.2939785`.

[18] G. Ke, Q. Meng, T. Finley, *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[20] M. Caron, H. Touvron, I. Misra, *et al.*, "Emerging properties in self-supervised vision transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9650–9660. DOI: `10.1109/ICCV48922.2021.00952`. [Online]. Available: `https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper.html`.