High-throughput microplastic screening system using multispectral fluorescence imaging under UV excitation

Asuka Sawamoto[a] and Shigeki Nakauchi[b]

[a,b]Department of Computer Science and Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi, Japan 441-8580

ABSTRACT

The widespread contamination of microplastics (MPs) necessitates comprehensive monitoring across various environments. However, standard analytical methods, such as FTIR and Raman spectroscopy, are time-consuming and resource-intensive, limiting their utility for high-throughput screening. Here, we propose a rapid multispectral imaging method using polymer auto-fluorescence under UV excitation. Leveraging the distinct fingerprints observed in Excitation-Emission Matrices (EEMs), our approach optimizes spectral acquisition by selecting a minimal subset of effective spectral bands rather than measuring the full sets of spectra in the EEM. Using a Random Forest-based feature selection, we demonstrated that only 8 spectral bands are sufficient to classify 9 representative virgin polymers with an average accuracy of 89%. This significant reduction in spectral dimensionality facilitates the design of simplified hardware using commercially available UV-LEDs, offering a practical and accessible solution for high-throughput MP monitoring.

Keywords: Microplastics, Multispectral fluorescence imaging, UV excitation, High-throughput screening, Spectral optimization, Excitation-emission matrix

1. INTRODUCTION

The widespread accumulation of microplastics (MPs) in the natural environment is an increasingly serious international issue [1,2]. Understanding MP pollution sources and dynamics requires comprehensively quantifying their distribution and flux across various media [3]. Specifically, tracking pollution sources and environmental behavior, identifying the specific material type, such as polyethylene (PE) from packaging vs. polyethylene terephthalate (PET) from bottles, of every single sample collected in field surveys is essential [4].

However, while conventional analytical methods such as Fourier Transform Infrared Spectroscopy (FTIR) and Raman spectroscopy excel at identifying the chemical structure of MPs, they rely on point-scanning, which requires an enormous amount of time for analysis [5, 6]. Consequently, comprehensive chemical profiling of the vast amounts of MPs in the environment is virtually impossible within a realistic timeframe. Therefore, for in-process applications, there is a strong demand for high-throughput solutions that achieve both practical classification accuracy and rapid analysis speeds [7].

In this study, we propose a multispectral imaging method specialized for polymer identification based on the Excitation-Emission Matrix (EEM). By utilizing inherent polymer auto-fluorescence under UV excitation, we capture essential spectral features as images. We first, toward the future construction of a high-throughput system, we conducted a feasibility study using a benchtop system equipped with a broadband light source. Furthermore, we identified a combination of spectral bands that minimizes measurement time while maintaining classification accuracy.

Further author information: (Send correspondence to A.S.)
A.S.: E-mail: sawamoto.asuka.us@tut.jp, Telephone: +81 90 7647 0330

2. PRINCIPLE

2.1 Fluorescence Characteristics

Most petroleum-based synthetic polymers emit specific auto-fluorescence under ultraviolet (UV) excitation due to their chemical structures, such as electron conjugated systems, and additives. This fluorescence characteristic depends on the Excitation-Emission Matrix (EEM) and serves as a unique fingerprint for each resin type. In this study, we utilize these EEM characteristics as the physical basis for MP classification. Generally, the fluorescence intensity F at an excitation wavelength λ_{ex} and an emission wavelength λ_{em} is described by the following simplified model:

$$F(\lambda_{ex}, \lambda_{em}) = K \cdot I_0(\lambda_{ex}) \cdot (1 - 10^{-\varepsilon(\lambda_{ex})cL}) \cdot \Phi(\lambda_{ex}, \lambda_{em}) \tag{1}$$

where K is the instrument constant (geometric factors and detection efficiency), I_0 is the excitation light intensity, ε is the molar absorption coefficient, c is the concentration, L is the optical path length, and Φ is the fluorescence quantum yield. Under the dilute solution or thin film approximation, Equation (1) becomes linear with respect to the material-intrinsic parameters ε and Φ. Since the spectral profiles of ε (absorption characteristics) and Φ (emission efficiency) differ depending on the polymer type, measuring these at multiple wavelengths enables material identification.

As a preliminary experiment, we measured the EEMs of the nine target polymers used in this study (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, ABS) using a fluorescence spectrophotometer (F-7000, Hitachi High-Tech). The obtained EEM spectra are shown in Figure 1. These results confirmed that each polymer possesses a specific fluorescence spectrum under UV excitation.

2.2 Data-Driven Band Selection

While high-dimensional data such as EEMs are essential for accurate polymer classification, acquiring and processing the full hyperspectral datacube is inefficient for real-time applications due to the massive data volume and spectral redundancy. Therefore, an approach is needed to selectively acquire only the minimal spectral bands that have a high contribution to discrimination. However, since spectral data possess strong correlations between bands and contain non-linear relationships, it is difficult to determine the optimal subset using simple linear models. In this study, we employed Random Forest (RF), which can account for non-linear interactions between features. In RF, the importance (Gini Importance) of a feature (wavelength band) X_j is defined as the sum of the decrease in impurity at the nodes where that feature was used for splitting, as shown in the following equation:

$$I_G(X_j) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_j} p(t) \Delta i(s_t, t) \tag{2}$$

where N_T is the total number of trees constituting the forest, $v(s_t)$ is the variable used for splitting at node t, $p(t)$ is the proportion of samples reaching node t, and $\Delta i(s_t, t)$ indicates the Gini Impurity Decrease due to the split. Based on this $I_G(X_j)$, we applied a step-wise method in which bands with high importance are sequentially added and evaluated. This allowed us to determine the optimal band configuration that minimizes imaging time while maintaining classification accuracy.

3. EXPERIMENTAL SETUP

3.1 Optical Configuration

To acquire comprehensive data for spectral optimization, we used a benchtop evaluation system equipped with a broadband light source [8]. The schematic diagram of the experimental apparatus is shown in Figure 2. We used a tunable light source (MAX-303, Asahi Spectra) equipped with a broadband xenon lamp. We selected the excitation wavelength via a rotating filter wheel and guided it to the sample through an optical fiber. The fluorescence emitted from the sample passed through an emission-side bandpass filter to remove scattered excitation light, and a UV-sensitive camera (BU-56DUV, Bitran) captured the signal.

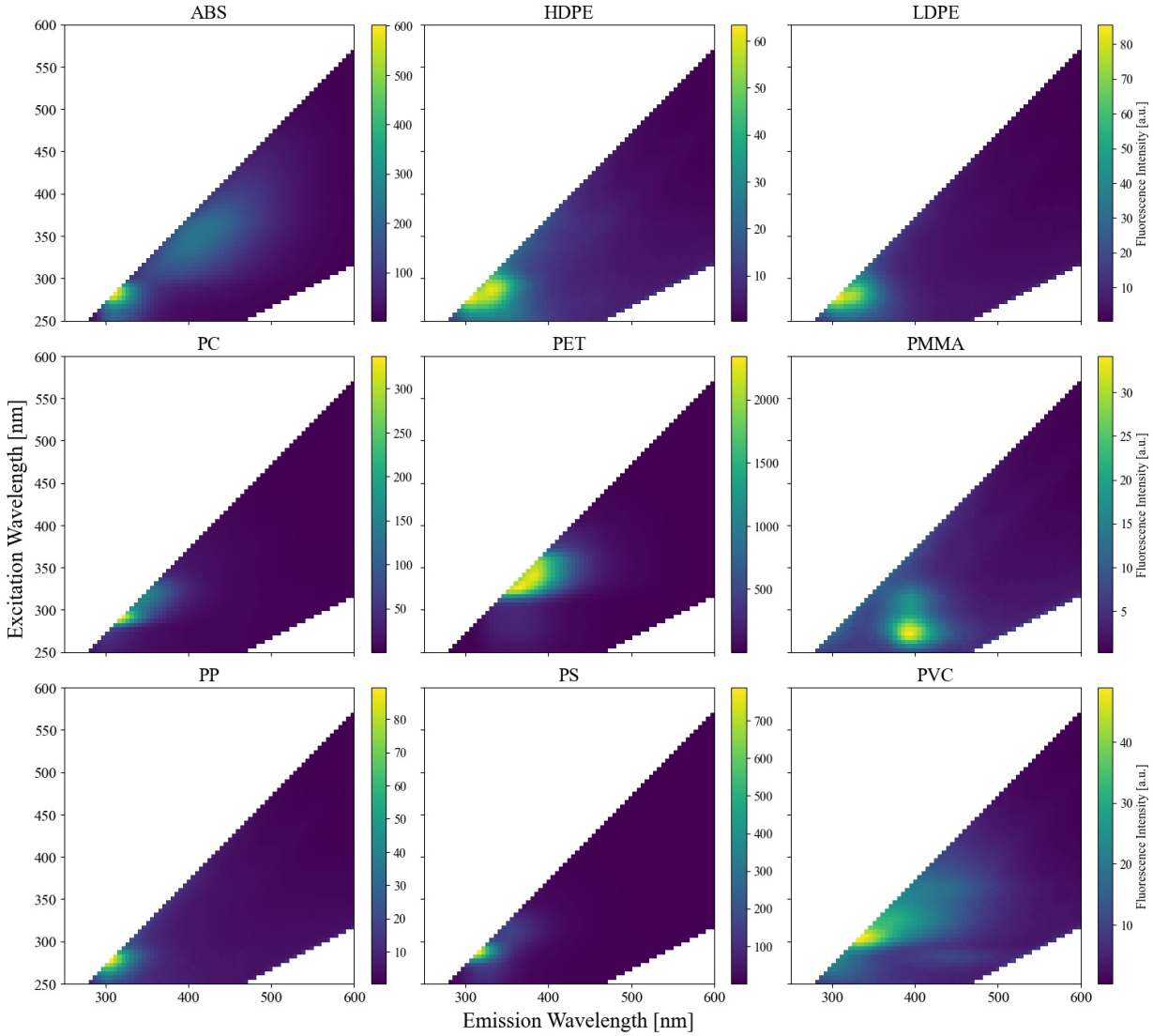


Figure 1: Excitation-Emission Matrices (EEMs) of the nine types of virgin polymer pellets used in this study. The contour plots represent the fluorescence intensity [a.u.], with resin names indicated above each plot. The distinct spectral patterns (fingerprints) observed for each polymer type demonstrate the feasibility of classification based on UV-excited fluorescence. Note: The diagonal regions corresponding to Rayleigh and Raman scattering have been computationally removed to enhance the visibility of fluorescence features.

To construct a comprehensive hyperspectral dataset, we combined excitation bandpass filters (center wavelengths: 260–380 nm) and emission bandpass filters (280–600 nm) at 20 nm intervals. By selecting valid combinations satisfying the Stokes shift (fluorescence wavelength > excitation wavelength), we acquired a total of 98 spectral images.

3.2 Samples and Data Acquisition

We used nine types of virgin resin pellets (HDPE, LDPE, PP, PS, PVC, PC, PMMA, PET, and ABS), which account for the majority of global plastic production, as experimental samples. The size of each pellet is approximately 3–5 mm (Figure 3a). While these are transparent or white under visible light and difficult to distinguish, they exhibit characteristic fluorescence intensity differences under UV excitation (Figure 3b, c). To improve data acquisition efficiency, we arranged the nine types of pellets within the same field of view during

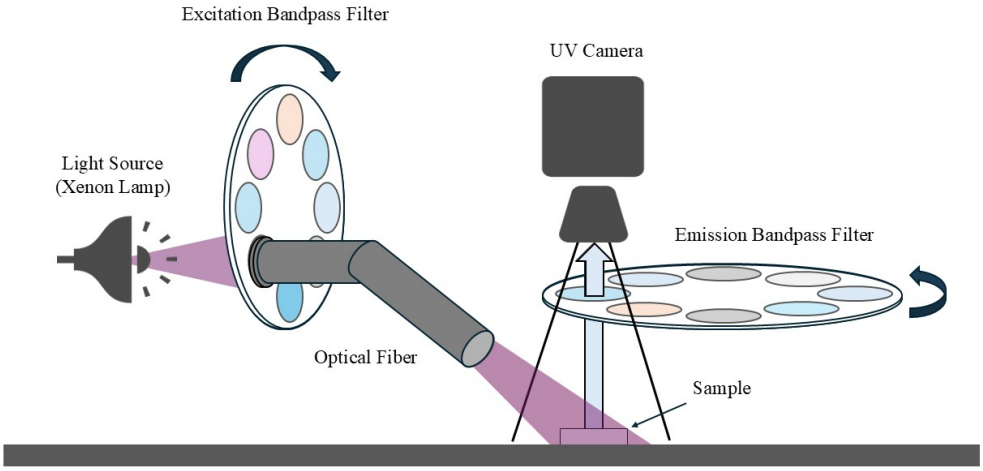


Figure 2: Schematic diagram of the benchtop evaluation system used for hyperspectral data acquisition. A broadband xenon lamp and interchangeable bandpass filters were used to excite the samples, and the emitted fluorescence was captured by a UV-sensitive camera to construct a comprehensive dataset for spectral optimization.

imaging. We acquired the datasets used for model calibration and evaluation in independent imaging sessions with different geometric arrangements:

- **Calibration Set:** To cover spectral variations depending on the arrangement state and construct a robust model, we integrated images acquired in the following two configurations:
  1. *Adjacent Arrangement:* Pellets of the same type are arranged to touch and cluster together.
  2. *Dispersed Arrangement:* Pellets are randomly dispersed so that they do not touch each other.
- **External Validation Set:** A dataset used solely for the final performance evaluation of the model, independent of the calibration process. It consists only of images with the *Dispersed Arrangement*. This verified the pure identification capability for unknown data.

4. METHOD

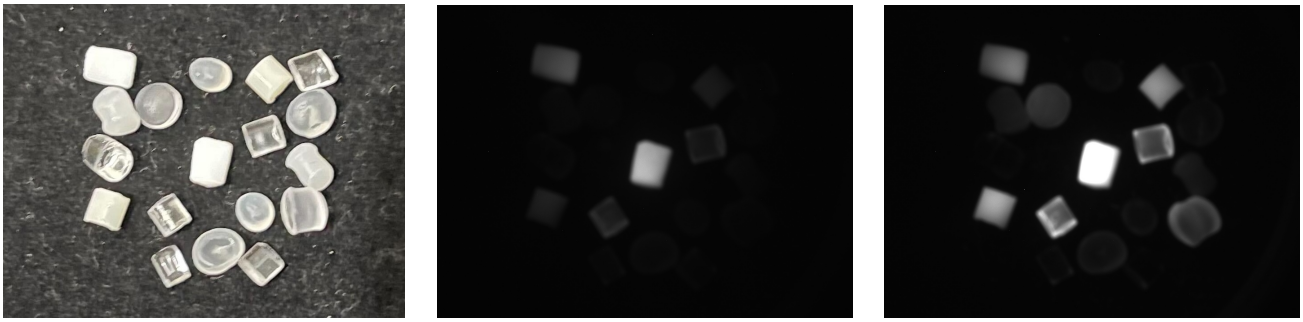
In this study, we propose a framework for high-throughput and accurate identification by selecting optimal wavelengths from acquired hyperspectral data. The process consists of three main stages: (1) Problem Formulation, (2) Pre-processing, and (3) Feature Selection.

4.1 Problem Formulation

Let the spectral intensity vector obtained from each pixel be $\mathbf{x} \in \mathbb{R}^D$ (where $D = 98$ corresponds to the number of acquired spectral images described in Section 3.1). Let the class label of the target polymer be $y \in \{C_1, \ldots, C_9\}$. The objective of this study is to learn a mapping function $f : \mathbf{x} \to y$ that predicts the correct label y from the input $\mathbf{x}$.

To verify the effectiveness of the proposed method, we define two experimental settings:

- **Baseline (Full-Spectrum):** A model that uses the entire wavelength band ($D = 98$) as input features. This serves as the theoretical upper limit of accuracy.
- **Ours (Optimized-Band):** A model that uses only k important bands selected by the algorithm described below ($k \ll 98$).



(a) Visible Light (b) Ex 320 / Em 400 (c) Ex 360 / Em 480

Figure 3: Images of the experimental samples consisting of nine types of virgin polymer pellets. (a) Visible light image; most samples appear transparent or white. (b, c) Representative fluorescence images acquired at (b) Ex 320 nm / Em 400 nm and (c) Ex 360 nm / Em 480 nm. Note: The sample arrangement differs between visible and fluorescence images. Despite the random arrangement, characteristic fluorescence intensity differences are clearly observed.

4.2 Pre-processing

The pixel intensity of the acquired fluorescence images strongly depends on the geometric shape of the sample, the distance from the light source, and the non-uniformity of illumination. These intensity variations hinder accurate polymer identification. Therefore, in this study, to achieve robust identification based solely on the polymer-specific "spectral profile" rather than absolute intensity, we applied Minimum Value Subtraction and L2 Normalization to the spectral data vector $\mathbf{x}$ of each pixel:

$$\mathbf{x}' = \frac{\mathbf{x} - \min(\mathbf{x})}{\|\mathbf{x} - \min(\mathbf{x})\|_2} \tag{3}$$

This process generates a normalized feature vector $\mathbf{x}'$ that is independent of lighting conditions and sample arrangement.

4.3 Feature Selection Strategy

To determine the optimal subset of spectral bands, we implemented the step-wise method based on the Random Forest Gini Importance described in Section 2.2. The specific procedure is as follows:

1. **Ranking (Importance Calculation):** We trained an initial RF model using the pre-processed calibration set and calculated the importance $I_G(X_j)$ of each band. The model input included all 98 available excitation-emission pairs.
2. **Iterative Evaluation:** We added features (bands) to the model one by one in descending order of calculated importance and evaluated discrimination performance at each step.
3. **Determination of Optimal Subset:** We adopted the "Weighted Average F1-Score of all 9 classes" as the evaluation metric to consider the balance of the entire system rather than just specific plastics. We selected the optimal band configuration at the point where the increase in F1-score saturated with the increase in the number of bands, and the number of images was minimized.

5. RESULTS

5.1 Optimization of Spectral Bands

The results of feature selection by the step-wise method using the calibration dataset are shown in Figure **??**. As the number of spectral bands k increased, the weighted F1-score for all 9 classes rose rapidly, exceeding 0.90 at $k = 2$. Subsequently, the improvement in the score became gradual, and we confirmed that the accuracy almost

saturated (above 0.99) around $k = 5$. Finally, considering the trade-off between slight accuracy improvement and computational cost, we adopted the band configuration of $k = 8$, which showed the highest stability, as the optimal solution. In this internal validation stage, the model demonstrated an extremely high precision rate.

5.2 Performance on External Validation Set

Using the selected 8 bands, we classified the external validation dataset to verify robustness against geometric variations. As a result, the model maintained a high weighted average F1-score (comparable to the calibration phase). As shown in Figure **??**, the model classified 7 out of the 9 types with extremely high precision (F1-score ≥ 0.98). We observed similar trends in the external data as in the calibration phase, which provides evidence that the constructed model generalizes well, rather than overfitting to specific sample arrangements or lighting conditions.

6. DISCUSSION

6.1 Interpretation of Optimization Behavior

In this study, we strategically prioritized global optimization to maximize the system's overall average performance suitable for simultaneous screening. Consequently, while we observed a temporary score decrease for specific classes (e.g., HDPE) due to class-specific trade-offs, the selected 8 bands achieved high generalization performance on the external validation set. This confirms that prioritizing the overall average over slight individual fluctuations is an effective and rational strategy for high-throughput applications.

6.2 High-Throughput Capability

The greatest advantage of this method is the overwhelming throughput achieved by area-scan imaging. While conventional point-scanning methods require time proportional to the number of measurement points, our method acquires spectral information for the entire field of view simultaneously. The experimental results demonstrate a speedup of several orders of magnitude compared to conventional methods.

6.3 Limitations and Potential Improvements

There are several limitations to this study.

First, we observed some confusion between PET and HDPE (PET misclassification rate: approx. 35%). We attribute the root cause of this misclassification to the significant feature overlap between these classes in the reduced 8-dimensional space, despite their distinct full spectral shapes.

While the current Random Forest model struggled to separate these overlapping clusters, introducing representation learning methods such as Deep Learning could potentially improve discrimination by learning more complex manifold structures.

Second is the scope of direct accuracy comparison. Since this study focuses on throughput improvement, we did not perform strict particle-to-particle comparison with conventional methods.

Third, we conducted this experiment using virgin pellets. Since microplastics in actual marine environments may have altered fluorescence characteristics due to biofilm adhesion or weathering, further verification using real environmental samples is necessary.

6.4 System Design for Field Deployment

We conducted this study as a proof-of-principle for the future construction of a high-throughput system. Based on the verification results, we designed a compact dedicated system for on-site implementation. Figure 4 shows the block diagram of the proposed system. To replace the large benchtop equipment, we adopted high-power UV-LEDs as the light source. However, commercially available LEDs have a wider Full Width at Half Maximum (FWHM) compared to bandpass filters, and the selectable center wavelengths are discrete. Therefore, in this design, we selected a group of commercially available LEDs that approximately cover the spectral regions suggested to be effective in the benchtop experiment. In the future, we plan to perform re-optimization of features based on actual hardware characteristics using this compact system. We will adopt a single-board computer (SBC, e.g., Raspberry Pi 5) for the control system, aiming to realize a low-cost, standalone monitoring solution.

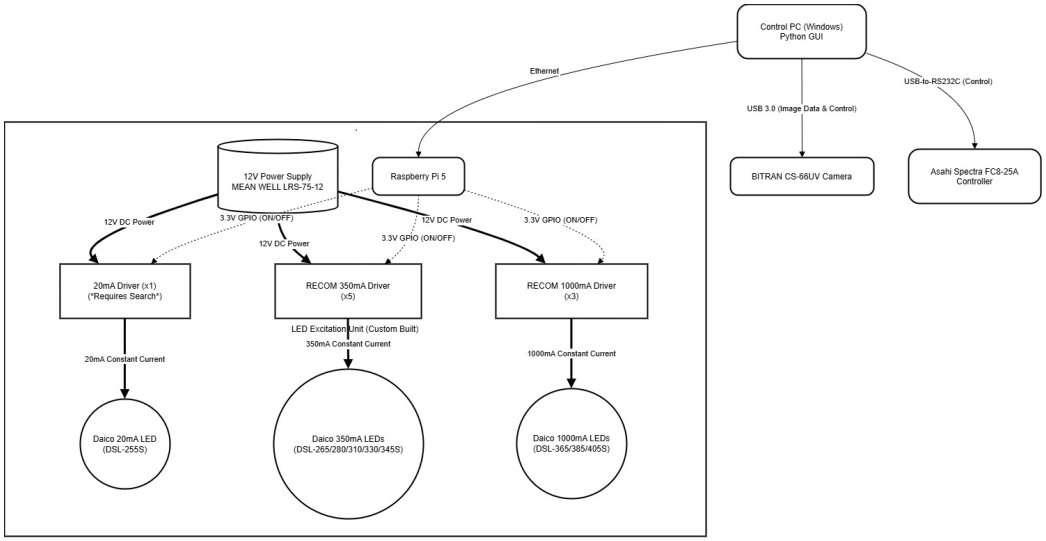


Figure 4: Block diagram of the proposed compact dedicated system. High-power UV-LEDs and an SBC are used to achieve a standalone, low-cost solution.

7. CONCLUSION

In this study, we demonstrated the feasibility of a high-throughput microplastic identification method using UV excitation fluorescence imaging. Through band selection using the step-wise method, we revealed that an average classification accuracy of 89% can be obtained with only 8 spectral bands. The proposed method achieves significantly higher throughput compared to conventional methods and demonstrates high industrial practicality. Based on these findings, we also proposed a design for a compact standalone system using UV-LEDs. In the future, we will focus on the construction of this dedicated hardware and demonstration experiments using real marine samples.

REFERENCES

[1] Hale, R. C., Seeley, M. E., La Guardia, M. J., Mai, L., and Zeng, E. Y., "A global perspective on microplastics," *J. Geophys. Res. Oceans* **125** (Jan. 2020).

[2] Jambeck, J. R., Geyer, R., Wilcox, C., Siegler, T. R., Perryman, M., Andrady, A., Narayan, R., and Law, K. L., "Marine pollution. plastic waste inputs from land into the ocean," *Science* **347**, 768–771 (Feb. 2015).

[3] Rocha-Santos, T. and Duarte, A. C., "A critical overview of the analytical approaches to the occurrence, the fate and the behavior of microplastics in the environment," *Trends Analyt. Chem.* **65**, 47–53 (Feb. 2015).

[4] Hidalgo-Ruz, V., Gutow, L., Thompson, R. C., and Thiel, M., "Microplastics in the marine environment: a review of the methods used for identification and quantification," *Environ. Sci. Technol.* **46**, 3060–3075 (Mar. 2012).

[5] Prata, J. C., da Costa, J. P., Duarte, A. C., and Rocha-Santos, T., "Methods for sampling and detection of microplastics in water and sediment: A critical review," *Trends Analyt. Chem.* **110**, 150–159 (Jan. 2019).

[6] Shim, W. J., Hong, S. H., and Eo, S. E., "Identification methods in microplastic analysis: a review," *Anal. Methods* **9**(9), 1384–1391 (2017).

[7] Serranti, S., Capobianco, G., Cucuzza, P., and Bonifazi, G., "Efficient microplastic identification by hyperspectral imaging: A comparative study of spatial resolutions, spectral ranges and classification models to define an optimal analytical protocol," *Sci. Total Environ.* **954**, 176630 (Dec. 2024).

[8] Bui, M. V., Rahman, M. M., Nakazawa, N., Okazaki, E., and Nakauchi, S., "Visualize the quality of frozen fish using fluorescence imaging aided with excitation-emission matrix," *Optics Express* **26**(18), 22954–22964 (2018).