Multimodal Image and Spectral Feature Learning for Efficient Analysis of Water-Suspended Particles

TOMOKO TAKAHASHI,[1,2,3,*] ZONGHUA LIU,[2,4] THANGAVEL THEVAR,[5] NICHOLAS BURNS,[5] DHUGAL LINDSAY,[6] JOHN WATSON,[5] SUMEET MAHAJAN,[3] SATORU YUKIOKA,[7] SHUHEI TANAKA,[7] YUKIKO NAGAI,[6] AND BLAIR THORNTON[2,8]

[1] *Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka, Kanagawa, 2370061, Japan*
[2] *Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro, Tokyo, 1538505, Japan*
[3] *Department of Chemistry and the Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK*
[4] *Ocean BioGeosciences group, National Oceanography Centre, European Way, Southampton SO14 3ZH, UK*
[5] *School of Engineering, University of Aberdeen, AB24 3UE, Scotland*
[6] *Institute for Extra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology, 2-15 Natsushima-cho, Yokosuka, Kanagawa, 2370061, Japan*
[7] *Graduate School of Global Environmental Studies, Kyoto University, Yoshida, Sakyo-Ku, Kyoto, 6068501, Japan*
[8] *Centre for In situ and Remote Intelligent Sensing, Faculty of Engineering and Physical Science, University of Southampton, Burgess Road, Southampton SO16 7QF, UK*
[*] *takahas@jamstec.go.jp*

Abstract: We have developed a method to combine morphological and chemical information for the accurate identification of different particle types using optical measurement techniques that require no sample preparation. A combined holographic imaging and Raman spectroscopy setup is used to gather data from six different types of marine particles suspended in a large volume of seawater. Unsupervised feature learning is performed on the images and the spectral data using convolutional and single layer autoencoders. The learned features are combined, where we demonstrate that non-linear dimensional reduction of the combined multimodal features can achieve a high clustering macro F1 score of 0.88, compared to a maximum of 0.61 when only image or spectral features are used. The method can be applied to long-term monitoring of particles in the ocean without the need for sample collection. In addition, it can be applied to data from different types of sensor measurements without significant modifications.

1. Introduction

In situ analysis of liquid-suspended particles has applications in environmental monitoring, healthcare and water quality control [1–3]. Particularly, monitoring of suspended particulate matters in the ocean requires the relative abundance of different particle types to be understood [4, 5]. Often these particles have sparse distributions (10 to several hundred particles/L) [6]. Non-destructive methods such as digital holography can image suspended particles in large volumes (~12 mL/s) of water with a high spatial resolution (~20 μm) without the need for any sample preparation [7–10]. Digital holographic cameras have been extensively used in marine monitoring to obtain information about particle size and shape [11–13], using machine learning techniques [14–16] to automatically identify different particle types. However, for particles like microplastics, morphological information alone is not sufficient to distinguish the different materials [17]. Knowledge of their chemical composition is important to understand the origin, route and consequences of environmental pollution [18]. Recently, the authors demonstrated holographic imaging and Raman spectroscopy for non-destructive analysis of water-suspended microplastic particle composition [19]. While the Raman spectroscopic analyzers previously used for *in situ* surveys observed back scattered lights from a target [20, 21], our setup observes forward scattered light and shows that both holographic imaging and Raman spectroscopic signals can be obtained from water-suspended particles using a single, compact optical setup. While the optical setup to perform combined imaging and spectroscopic measurements of particles has been demonstrated, it is also necessary to develop analytical methods that can efficiently process multimodal data in order to take full advantage of such a setup. For multimodal data fusion analysis, the audio-visual emotion challenge to develop machine learning methods for automatic audio, visual and audiovisual emotion analysis is a well-known topic [22]. Similar to how human beings naturally process multimodal information [23], a number of publications have reported improvement of the recognition accuracy of emotions by multimodal fusion analysis of speech data (*e.g.* vocal effect) and visual data (*e.g.* face expression) from unimodal analysis [24–26]. In addition, novel multimodal deep-learning based methods have been demonstrated to further increase the accuracy [27, 28]. Data fusion applications have been expanded to a wide range of multi-sensory data analysis [29], such as biomedical diagnostics [30, 31], pharmacy [32, 33], automatic robot navigation [34], and remote sensing [35]. However, the previous methods have not been applied for the identification of marine particle types/materials due to the limitation of multiple sensory applications to analyze particles.

In this paper, we demonstrate the automatic clustering and classification of different types of marine particles by applying a simple data fusion technique to morphological (*i.e.* holographic images) and chemical (*i.e.* Raman spectra) data. We propose a multimodal learning method using autoencoders and further t-SNE dimensionality reduction, and compare the classification accuracy between uni and multimodal data with and without t-SNE. We investigate how unsupervised feature learning methods can be used to automatically extract and further combine multimodal features from different types of sensor measurements, and use these to efficiently identify different particle types.

2. Experiments

2.1. Samples

Experiments were performed on plankton, foraminifera, minerals and microplastic particles, where these were chosen based on their relevance to climate change and pollution monitoring [1, 36]. These were measured in artificial seawater, which is often used for method validation for marine sensing applications [37–39], to minimize the effect of water quality fluctuation on images and spectra. Plankton absorbs around 50 billion tons of carbon each year, accounting for 40 % of atmospheric CO_2 removal [40, 41]. Removed carbon is either stored as organic carbon as in the case of the copepods used in our experiments, which are one of the most abundant zooplankton species in the ocean, or as inorganic carbon as in the case of foraminifera, a single-cell organism with an external shell made of calcium carbonate. Our experiments also study sphalerite rock fragments, which are a common sulfide mineral in ores. The ability to monitor sulfide particle distributions is important for studying the potential impacts of sub-sea mining [42]. Finally, we investigate polypropylene (PP) and polyethylene (PE) microplastic pre-production plastic pellets (nurdles). PP and PE are selected since these are the most common types of microplastics found in aquatic environments [43]. We also investigate PE fragments that were collected from the ocean. The particle types and sample numbers for each type are summarized in Table 1.

Copepods were collected from the surface seawater during the KM20-11 cruise of the research vessel (R/V) Kaimei in December 2020 and kept in a freezer to preserve their morphological characteristics. The samples were defrosted using lukewarm water before the measurement.

Table 1. **Samples used in experiments**

Particle Type	Description	Number of samples
Organic carbon	Copepod	3
Inorganic carbon	Foraminifera	3
Mineral	Sphalerite	3
Microplastics	PP (nurdle)	3
Microplastics	PE (nurdle)	3
Microplastics	PE (marine)	3

Dried foraminifera samples (*Calcarina gaudichaudii*) were collected from Okinawa, Japan. The sphalerite rock fragments were collected from Daikoku Ore in Saitama, Japan. PP and PE nurdles were provided by Daikei Chemical, Inc. PE fragments were recovered from the surface seawater in Osaka Bay, Japan in September, 2018. These samples were separated from other particles by first dissolving biotic organic matter and performing Fourier transform infrared spectroscopy on the dried residue to identify the PE fragments. All particles used in our experiments had a dimension between 1 and 5 mm, and 3 different samples of each particle type were measured to assess the performance of our method.

2.2. Setup

The integrated in-line holographic imaging and Raman spectroscopy setup used in our experiments is shown in Fig. 1 and has previously been described in Ref. [19]. A quartz glass cell of length 20 cm and diameter 20 mm (Sterna cell, 34-Q-200) was filled with artificial seawater and illuminated by a collimated laser of 10 mm beam diameter. A single longitudinal mode continuous wave (CW) laser (Oxxius, LCX-532S-300) beam with a wavelength of 532 nm was delivered via a single-mode fiber. The exiting beam from the fiber was collimated and passed through a bandpass filter (Semrock, LL01-532-25) before entering the measurement cell. The laser power was set at 160 mW at the output of the bandpass filter. After passing through the measurement cell, the beam was split using a 532 nm dichroic beam splitter (Semrock, Di03-R532-t1-25x36). The reflected beam was used for holographic imaging. It passed through an attenuation filter (Sigma Koki, MFND-25-0.1) before a hologram was recorded by a two-dimensional complementary metal-oxide semiconductor (CMOS) 2464 $\times$ 2056 pixel array (JAI, GO-5100-USB). Images were taken continuously with a 50 μs exposure time. The lights with wavelengths longer than 532 nm were transmitted through the beam splitter and collected for Raman spectroscopy via a set of lenses (Thorlabs, F810SMA-543) that was mounted to a multi-mode fiber (Thorlabs, M29L01). A 532 nm longpass filter (Semrock, BLP01-532R-25) was placed before the fiber to ensure blocking of the 532 nm beam. A spectrometer with a wavenumber range from 200 to 3100 cm^{-1} and a resolution of 10 cm^{-1} (Wasatch Photonics, WP-532-A-S-ER-10) was used. The acquisition period was set at 5 s to maximize signal to noise ratio while avoiding saturation.

2.3. Data acquisition

The holographic imaging detector records the interference patterns generated from the interaction between the unscattered laser beam (reference beam) and the scattered light by the particles (object beam). To recover information on particle morphology, the interference patterns are reconstructed as described previously by the authors [10, 44, 45], using the angular spectrum method [46, 47]. Copepods, foraminifera, and mineral particles immediately sank to the bottom

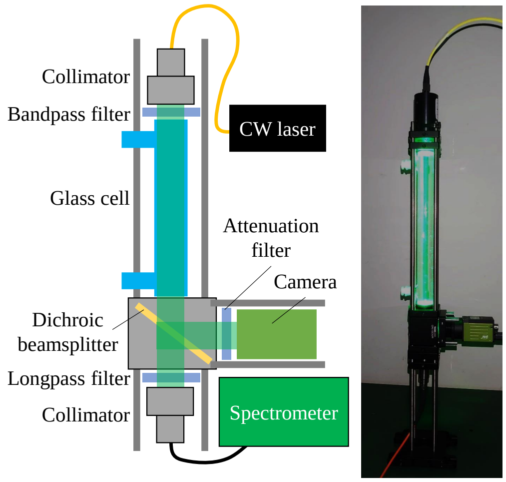

Fig. 1. Experimental setup. A 532 nm single longitudinal mode laser is used to illuminate samples suspended in bulk artificial seawater. A beam splitter is used to take holographic images and Raman spectra using the same setup with different exposure times.

of the measurement cell while the plastics floated due to their buoyancy. Therefore, the relative distances between the samples, laser and detector were consistent for each particle type. Fig. 2 (a) shows examples of bright field microscopic images of the samples. Fig. 2 (b) shows the corresponding reconstructed holographic images of the seawater-immersed particles that were measured using the experiment setup. Morphological characteristics unique to copepods (*i.e.* antennae and legs) and foraminiferas (*i.e.* spines) are clearly seen in the holographic images, whereas other particles are not obviously distinguished. 100 holographic images of each sample were taken, where the measurement cell was shaken and rotated between images so that the samples were imaged from different angles and directions. The width of the images was trimmed to 2056 pixels so as to cut off the unilluminated region, and it was manually confirmed that the whole sample was visible in all images. The images were normalized so that each image's maximum and minimum pixel intensities were 1 and 0, respectively.

120 Raman spectra were taken for each sample. To reduce noise, 50 spectra were randomly selected and averaged, where this process was repeated using the boot-strapping method [48] to produce 100 unique spectra [37]. The background spectrum was taken using the same setup without any target particles and the signal was averaged in the same way. Each averaged spectrum was normalized by setting the S-O stretching peak at 981 cm^{-1} to have unitary intensity. This peak was chosen as it is always present in seawater due to dissolved $SO_4{}^{2-}$ [49]. The background spectrum was subtracted from the averaged spectrum for each particle sample to remove the contributions of the optical setup and seawater. The spectral range from 300 to 1711 cm^{-1} (309 pixels) was used for analysis since the wavenumbers out of this range do not have many Raman peaks. Fluorescence signals were modeled in the range and subtracted using an eighth or ninth-order polynomial asymmetric truncated quadratic function depending on the samples. The most suitable order was experimentally determined, using the MATLAB^{TM} "backcor" function [50], which estimates background signals by minimizing a non-quadratic cost function. Fig. 2 (c) shows examples of processed Raman spectra for each sample type. Strong Raman peaks of PP and PE (PP: 809, 841, 1152, 1167, 1330, and 1458 cm^{-1}, PE: 1062, 1130, 1170, 1295, 1418, 1440, 1461 cm^{-1} [51]) are observed in the spectra of nurdles as these samples are semi-transparent, enabling high efficiency collection of forward Raman scattering, while for

other particles the Raman peaks are generally less distinct, due to high opacity of the targets. Peaks at 1062, 1295, and 1440 cm^{-1} are observed in the spectra of PE fragments, although peaks are not as strong as the ones seen in PE nurdle spectra due to the interference from green pigments. An intense band from carotenoid is seen at 1521 cm^{-1} [52] in copepod spectra. A peak assigned to the symmetric stretching vibration of the CO_3^{2-} ion is seen at 1090 cm^{-1} [53] in the foraminifera spectra, while other unidentified peaks are also observed. The overall intensities of mineral spectra are weaker than other spectra with no strong peaks observed.

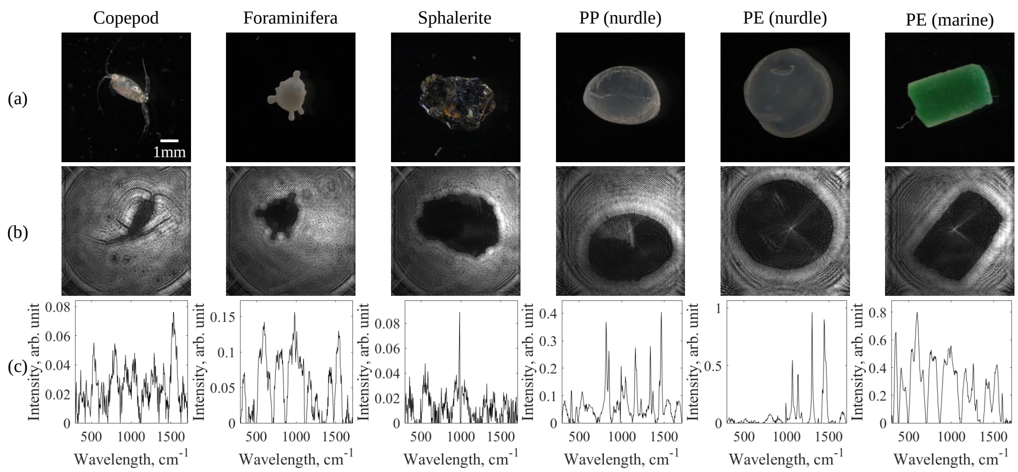

Fig. 2. Examples of (a) bright field microscopic images, (b) reconstructed holographic images and (c) processed Raman spectra for each particle type.

2.4. Unsupervised feature learning

We investigated autoencoder-based unsupervised feature learning approaches to group the different particle types. The advantage of unsupervised methods is that they do not rely on human-labeled data for training, which do not always exist and are often time consuming to generate [54]. Autoencoders are a generic type of unsupervised feature learner that has been well established for the analysis of imagery, including holographic images [55]. They consist of an encoder network, which reduces the input data down to smaller latent representations, and a decoder network that attempts to reconstruct the original data from the compressed latent representation. The latent representations through optimization of both networks to minimize the difference between the original inputs and their reconstructions can be used as features for clustering and classification tasks [56]. Classification based on features extracted using autoencoders can outperform the use of features that have traceable physical meaning such as principal component analysis [57, 58]. A key advantage is that they are unsupervised, and can flexibly manage different sizes and dimensionality of data inputs as well as the size of the latent feature space representations they output, without significant modification of their underlying form, which is suitable for multimodal data [29]. Fig. 3 illustrates the proposed multimodal holographic image and Raman spectrum feature learning. A convolutional autoencoder is used to extract features from the holographic image reconstructions. Deep-learning convolutional autoencoders based on Alexnet have been successfully developed for sub-sea image classification [59, 60]. When applied to holographic images, improvement of clustering performance was found when a modified AlexNet where the fully-connected layers were replaced by two convolution layers was used [45]. Here we use the same modified AlexNet-based deep learning autoencoder described in Ref. [45], which was well tuned for in-line holographic images. The entire dataset (1800 images) was used to train the network after reducing each image to 227 × 227 pixels to fit the input layer. When only images were used in the subsequent analysis, 16 latent features were extracted based on recommendations of prior work [59]. This was reduced to 8 when features were combined with those extracted from spectra so that the total number of extracted features was maintained. Information about the particle type was only used for performance validation, and was not used in training. The Raman spectra obtained with our setup are one-dimensional (309 × 1) and have a significantly smaller data size than the holographic images. A single-layer autoencoder was used to learn features where the latent representation size was set to 16 when only spectral information was used, and to 8 when features were combined with those extracted from holographic images.

Once features are extracted from the encoders, k-means clustering is used to group particles. This method was chosen as it is unsupervised and so does not require any human-labeled training data. We note that while different unsupervised clustering approaches such as random forest and self-organized maps, or supervised methods such as support vector machines, neural network classifiers or Gaussian processes may improve overall scores, the focus of this paper is on improving the quality of the features used for subsequent analysis, and such optimization of clustering or classification methods is beyond our scope.

The number of clusters was set to 6, which equals the number of particle types used in this study. We investigated two grouping methods. The first method is feature-level fusion, and directly uses the latent representations. The second method is model-level fusion and uses non-linear dimensional reduction to further compress the latent representations prior to clustering. For the direct approach, k-means clustering is carried out directly on the features extracted from holographic images (condition D1), Raman spectral data (condition D2), and on the combined features (condition D3), respectively. The latent space was set so that the final number of features used for clustering was the same, at 16 features, across all experimental conditions to allow for a fair comparison. For the reduced approach, a further reduction from 16 to 2 dimensions is achieved using the non-linear t-distributed stochastic neighbor embedding (t-SNE) [57, 61]. Clustering is performed on the reduced two-dimensional features extracted from holographic images (condition R1), Raman spectral data (condition R2), and on the combined features (condition R3), respectively. Clustering performance is assessed using confusion matrices and F1-average score (*i.e.* macro F1 score [62]), where cluster to particle type correspondence is achieved by determining the largest number of particles of a given type falling within each cluster. The different experimental conditions investigated in this work are summarized in Table 2.

Table 2. **Experimental conditions analyzed in this work.**

	D1	D2	D3	R1	R2	R3
Images features	16	0	8	16	0	8
Spectral features	0	16	8	0	16	8
Dimension reduction				✓	✓	✓
Total features	16	16	16	2	2	2
Clusters k	6	6	6	6	6	6

3. Results and discussion

Fig. 4 shows the t-SNE plots of the latent representations extracted from (a) holographic images, (b) Raman spectroscopy, and (c) their combination. The color of data points indicates particle type (black: copepod, red: foraminifera, blue: mineral, pink: PP nurdle, purple: PE nurdle,

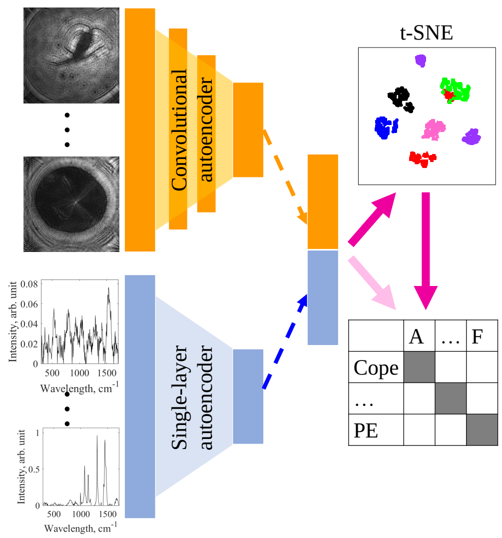

Fig. 3. Diagram of processes for combining features extracted from holographic image and Raman spectra, which are used for clustering either directly or after applying t-SNE dimensional reduction.

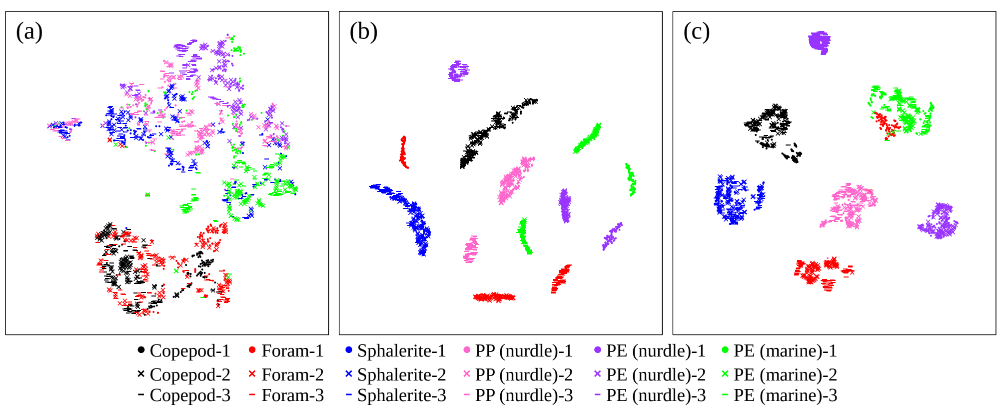

Fig. 4. t-SNE visualization of latent representations extracted from (a) holographic images, (b) Raman spectra, and (c) combined. The shape in the legend indicates three different samples among the same type of particles (circle, cross, bar).

green: PE fragment). Table 3 shows the confusion matrix result of *k*-means clustering applied directly to the extracted features, and Table 4 shows the result of clustering applied to the extracted features that have been further reduced using t-SNE. The clustering groups A-F were automatically allocated to six clusters with the combination which gives the best F1-average score. Table 5 shows the F1-scores for each particle type and processing condition.

Using features extracted from holographic images alone (D1, R1), it can be seen that copepods and foraminiferas form one mixed cluster. The remaining four particle types form the second cluster. This can be understood by looking at the examples in Fig. 2, where copepods and foraminifera have complex shapes, while the remaining particle types have a simpler form. PE fragments have an angular shape that distinguishes them from the round shape of the mineral and PE, PP nurdles, where this pattern can be seen by the increased separation between it and the other particle types. Clustering with $k = 6$ results in groupings with mixed particle types, where an overall trend that two clusters dominate is reflected in the confusion matrices for D1 (Table 3 (a)) and R1 (Table 4 (a)). The F1-score averages are higher for clustering after using t-SNE for dimensional reduction rather than direct use of the latent representations.

For Raman spectral data (D2, R2), 13 distinct groupings can be seen, where for most particle types the individual samples are separated. While copepods and minerals form their own groups for all samples, other particle types form two or three separate clusters for each type, which are not necessarily close together in the latent representation space. This reflects the sensitivity of Raman spectroscopy-based features to differences in the individual samples regardless of particle type. The over discrimination is seen in the confusion matrices for D2 and R2 in Tables 3 (b) and 4 (b), respectively. The individual samples fall in or out of the six clusters in a binary manner, where the precision and recall rates for direct use of extracted features vary from 0 to 100 %. Although this trend is improved after t-SNE, the overall accuracy according to F1 scores is reduced, where dimensional reduction results in poorer accuracy for the plastic particles in particular. The results show that it is not possible to reliably cluster features from Raman spectra to map onto the 6 particle types. The average F1 scores for holographic images (D1) and Raman spectra (D2) have similar values of around 0.5 and 0.6, respectively, where further dimensional reduction improves the score for holographic images (R1), but not for Raman spectra (R2).

Combining the features from holographic images and Raman spectra improves the F1 scores for both the direct (D3) and the reduced t-SNE based (R3) clustering. In particular, dimensional reduction results in significant performance gains where both data types are combined. This is seen with foraminifera, where direct use of the latent representations has poor precision and recall, but dimensional reduction improves these from 3 % to 97 % and 2 % to 66 %, respectively. D3 and R3 confusion matrices are shown in Table 3 (c) and Table 4 (c), respectively.

Table 5 shows that combining features gives the highest F1 score for all particle types investigated in this work. The highest average F1 score of 0.88, is obtained for condition R3, where combined features after non-linear dimensional reduction using t-SNE are used. This score is 0.25 higher than for the directly combined case, and ≥0.27 higher than when holographic images or Raman spectra based features are used in isolation. Condition D3 gives the second best results. For condition R3, all particle types have F1 values over 0.79, demonstrating reliable mapping of the clusters onto the particle types of interest. The large performance gain when non-linear dimensional reduction is applied to the combined features can make effective use of the favorable characteristics of each measurement type. The t-SNE plot in Fig. 4 (c) shows that copepods, minerals, and PP nurdles form groups with well separated boundaries. One sample of PE nurdles forms a group that is independent of others and one sample of foraminifera merges with a cluster of PE fragments. In both cases, it could be assumed to be mainly due to the features of Raman spectra as these trends are also seen in the t-SNE visualization of Raman spectral latent representations (Fig. 4 (b)). This could be mitigated by using fewer features of Raman spectra. In future works aiming at real-sea applications, fine tuning of models including selecting the best combination of the number of features among different data types will be performed to improve clustering and classification performances.

The results show that features extracted using an appropriately designed autoencoder and further use of t-SNE for non-linear dimensional reduction significantly improves the quality of the features available to describe different particle types, and this improvement enhances classification accuracy. For application to *in situ* monitoring of marine particles, the method needs to be verified on larger numbers and types of particles to be more representative of the variety of morphological and compositional combinations that exist in nature. However, the study has demonstrated a novel approach to combine features learned from multiple different sensing modes, which improves clustering performance for a diverse range of marine particle types. Since the proposed method of combining and blending features can be applied to any

Table 3. **Confusion matrix between particle type and the clustering result created using k-means for (a) holographic images D1, (b) Raman spectra D2, and (c) combined D3 latent representations. A-F indicate clustering groups.**

(a) D1

	A	B	C	D	E	F	Recall
Copepod	225	75	0	0	0	0	75 %
Foram	133	163	0	0	0	4	54 %
Sphalerite	53	25	126	30	17	49	42 %
PP (nurdle)	1	4	108	93	39	55	31 %
PE (nurdle)	0	1	84	25	140	50	47 %
PE (marine)	8	63	17	71	59	82	27 %
Precision	54 %	49 %	38 %	42 %	55 %	34 %	

(b) D2

	A	B	C	D	E	F	Recall
Copepod	300	0	0	0	0	0	100 %
Foram	0	100	0	100	0	100	33 %
Sphalerite	300	0	0	0	0	0	0 %
PP (nurdle)	0	0	0	300	0	0	100 %
PE (nurdle)	0	0	100	0	200	0	67 %
PE (marine)	0	0	0	0	0	300	100 %
Precision	50 %	100 %	0 %	75 %	100 %	75 %	

(c) D3

	A	B	C	D	E	F	Recall
Copepod	296	1	0	0	0	3	99 %
Foram	176	7	0	1	0	116	2 %
Sphalerite	4	76	192	6	21	1	64 %
PP (nurdle)	0	54	40	199	4	3	66 %
PE (nurdle)	0	49	18	4	229	0	76 %
PE (marine)	2	16	3	32	0	247	82 %
Precision	62 %	3 %	76 %	82 %	90 %	67 %	

input data type using encoded latent representation spaces, the method forms a versatile approach to combine measurements taken from multiple sensors with different data types and sizes, and makes efficient use of the favorable characteristics of each measurement type.

4. Conclusion

We have proposed a novel method to combine features extracted from images and spectra of seawater-suspended particles. Features were first extracted from data taken of the same target

Table 4. **Confusion matrix between particle type and the clustering result created using k-means after t-SNE dimensional reduction for (a) holographic images R1, (b) Raman spectra R2, and (c) combined R3 latent representations. A-F indicate clustering groups.**

(a) R1

	A	B	C	D	E	F	Recall
Copepod	220	80	0	0	0	0	73 %
Foram	146	147	6	0	0	1	49 %
Sphalerite	1	0	184	26	22	67	61 %
PP (nurdle)	0	0	121	120	58	1	40 %
PE (nurdle)	0	0	38	102	158	2	53 %
PE (marine)	0	29	27	50	35	159	53 %
Precision	60 %	57 %	49 %	40 %	58 %	69 %	

(b) R2

	A	B	C	D	E	F	Recall
Copepod	300	0	0	0	0	0	100 %
Foram	0	200	100	0	0	0	67 %
Sphalerite	0	0	142	158	0	0	47 %
PP (nurdle)	0	0	0	100	200	0	33 %
PE (nurdle)	100	0	0	0	100	100	33 %
PE (marine)	0	42	0	0	61	197	66 %
Precision	75 %	83 %	59 %	39 %	28 %	66 %	

(c) R3

	A	B	C	D	E	F	Recall
Copepod	300	0	0	0	0	0	100 %
Foram	0	199	0	1	0	100	66 %
Sphalerite	0	0	300	0	0	0	100 %
PP (nurdle)	0	7	0	293	0	0	98 %
PE (nurdle)	100	0	0	0	200	0	67 %
PE (marine)	0	0	0	0	0	300	100 %
Precision	75 %	97 %	100 %	100 %	100 %	75 %	

using an integrated setup for holographic imaging and Raman spectroscopy. Convolutional and single-layer autoencoders were used for holographic images and Raman spectra, respectively. While combining latent representations (feature-level fusion) slightly enhanced the macro F1 average score, the performance is further significantly improved by performing non-linear dimensional reduction (model-level fusion) using t-SNE on the combined latent representations. This increases the calculated accuracy from 0.63 to 0.88 using t-SNE, and the use of combined

Table 5. **Comparison of F1 scores, where the highest scores for each particle type are in bold.**

	D1	D2	D3	R1	R2	R3
		w/o t-SNE			w/ t-SNE	
	Holo	Raman	Fusion	Holo	Raman	Fusion
Copepod	0.63	0.67	0.76	0.66	**0.86**	**0.86**
Foram	0.52	0.5	0.03	0.53	0.74	**0.79**
Sphalerite	0.40	0	0.69	0.54	0.52	**1**
PP (nurdle)	0.36	0.86	0.73	0.40	0.36	**0.99**
PE (nurdle)	0.51	0.8	**0.83**	0.55	0.30	0.8
PE (marine)	0.30	0.85	0.74	0.6	0.66	**0.86**
Average	0.45	0.61	0.63	0.55	0.57	**0.88**

features outperformed a single information source for all particle types studied in this work.

Although our experiments used holographic images and Raman spectroscopy, the proposed method can be adapted to other types of sensor measurements. The use of convolutional and conventional autoencoders can learn and extract features from any two- or one- dimensional data type (*e.g.* images, spectra) without the need for labeled training datasets, respectively. Since dimensional reduction is performed on the feature space, it can efficiently combine features derived from other sensing methods and be applied to other measurement targets with minimal modification.

Funding. Japan Science and Technology Agency SICORP and Natural Environment Research Council (JST-NERC SICORP Marine Sensor Proof of Concept Grant JPMJSC1705, NE/R01227X/1); JSPS KAKENHI Grant (18K13934 and 18H03810); Sumitomo Foundation: Grant for environmental Research Project (203122).

Acknowledgments. The authors thank Dr. T. Fukuba for the support for building the experimental setup. The authors also thank Dr. H. Sawada for providing samples for this work.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. International Ocean Carbon Coordination Project, "Essential Ocean Variable (EOV): Particulate Matter," Retrieved June 20, 2022 (2017). Https://www.goosocean.org/eov.
2. T. Sun and H. Morgan, "Single-cell microfluidic impedance cytometry: a review," Microfluid. Nanofluidics **8**, 423–443 (2010).
3. V. Gauthier, B. Barbeau, R. Millette, J.-C. Block, and M. Prevost, "Suspended particles in the drinking water of two distribution systems," Water Sci. Technol. Water Supply **1**, 237–245 (2001).
4. E. Boss, L. Guidi, M. J. Richardson, L. Stemmann, W. Gardner, J. K. Bishop, R. F. Anderson, and R. M. Sherrell, "Optical techniques for remote and in-situ characterization of particles pertinent to geotraces," Prog. Oceanogr. **133**, 43–54 (2015).
5. A. M. McDonnell, P. J. Lam, C. H. Lamborg, K. O. Buesseler, R. Sanders, J. S. Riley, C. Marsay, H. E. Smith, E. C. Sargent, R. S. Lampitt *et al.*, "The oceanographic toolbox for the collection of sinking and suspended marine particles," Prog. oceanography **133**, 17–31 (2015).
6. D. J. Lindsay, A. Yamaguchi, M. M. Grossmann, J. Nishikawa, A. Sabates, V. Fuentes, M. Hall, K. Sunahara, and H. Yamamoto, "Vertical profiles of marine particulates: a step towards global scale comparisons using an autonomous visual plankton recorder," Bull. Plankton Soc. Jpn. **61**, 72–81 (2014).

7. J. Watson and P. W. Britton, “Preliminary results on underwater holography,” Opt. Laser Technol. pp. 215–216 (1983).
8. K. L. Carder, “A holographic micro-velocimeter for use in studying ocean particle dynamics,” in *SPIE 0160, Ocean Optics V,* vol. 160 (1978), pp. 63–66.
9. K. L. Carder, R. G. Steward, and P. R. Betzer, “In situ holographic measurements of the sizes and settling rates of oceanic particulates,” J. Geophys. Res. **87**, 5681–5685 (1982).
10. Z. Liu, T. Takahashi, D. Lindsay, T. Thevar, M. Sangekar, H. K. Watanabe, N. Burns, J. Watson, and B. Thornton, “Digital in-line holography for large-volume analysis of vertical motion of microscale marine plankton and other particles,” IEEE J. Ocean. Eng. **46**, 1248–1260 (2021).
11. J. Watson, S. Alexander, G. Craig, D. C. Hendry, P. R. Hobson, R. S. Lampitt, J. M. Marteau, H. Nareid, M. A. Player, K. Saw, and K. Tipping, “Simultaneous in-line and off-axis subsea holographic recording of plankton and other marine particles,” Meas. Sci. Technol. **12** (2001).
12. J. Watson, S. Alexander, V. Chavidan, G. Craig, A. Diard, G. L. Foresti, S. Gentili, D. C. Hendry, P. R. Hobson, R. S. Lampitt, H. Nareid, J. J. Nebrensky, A. Pescetto, G. G. Pieroni, M. A. Player, K. Saw, S. Serpico, K. Tipping, and A. Trucco, “A holographic system for subsea recording and analysis of plankton and other marine particles (HOLOMAR),” in *Oceans Conference (IEEE),* (2003), pp. 830–837.
13. R. B. Owen and A. A. Zozulya, “In-line digital holographic sensor for monitoring and characterizing marine particulates,” Opt. Eng. **39**, 2187 (2000).
14. V. Bianco, P. Memmolo, P. Carcagnì, F. Merola, M. Paturzo, C. Distante, and P. Ferraro, “Microplastic identification via holographic imaging and machine learning,” Adv. Intell. Syst. **2**, 1900153 (2020).
15. Y. Zhu, C. H. Yeung, and E. Y. Lam, “Microplastic pollution monitoring with holographic classification and deep learning,” J. Physics: Photonics **3**, 024013 (2021).
16. L. MacNeil, S. Missan, J. Luo, T. Trappenberg, and J. LaRoche, “Plankton classification with high-throughput submersible holographic microscopy and transfer learning,” BMC ecology evolution **21**, 1–11 (2021).
17. V. Bianco, D. Pirone, P. Memmolo, F. Merola, and P. Ferraro, “Identification of microplastics based on the fractal properties of their holographic fingerprint,” ACS Photonics **8**, 2148–2157 (2021).
18. V. Hidalgo-Ruz, L. Gutow, R. C. Thompson, and M. Thiel, “Microplastics in the marine environment: A review of the methods used for identification and quantification,” Environ. Sci. & Technol. **46**, 3060–3075 (2012).
19. T. Takahashi, Z. Liu, T. Thevar, N. Burns, S. Mahajan, D. Lindsay, J. Watson, and B. Thornton, “Identification of microplastics in a large water volume by integrated holography and raman spectroscopy,” Appl. Opt. **59**, 5073–5078 (2020).
20. X. Zhang, W. J. Kirkwood, P. M. Walz, E. T. Peltzer, and P. G. Brewer, “A review of advances in deep-ocean Raman spectroscopy,” Appl. Spectrosc. **66**, 237–49 (2012).
21. J. A. Breier, C. R. German, and S. N. White, “Mineral phase analysis of deep-sea hydrothermal particulates by a Raman spectroscopy expert algorithm: Toward autonomous in situ experimentation and exploration,” Geochem. Geophys. Geosystems **10**, 1–12 (2009).
22. B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, “Avec 2011–the first international audio/visual emotion challenge,” in *International Conference on Affective Computing and Intelligent Interaction,* (Springer, 2011), pp. 415–424.
23. S. Shimojo and L. Shams, “Sensory modalities are not separate modalities: plasticity and interactions,” Curr. opinion neurobiology **11**, 505–509 (2001).
24. S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: From unimodal analysis to multimodal fusion,” Inf. Fusion **37**, 98–125 (2017).
25. S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian, “Learning affective features with a hybrid deep model for audio–visual emotion recognition,” IEEE Transactions on Circuits Syst. for Video Technol. **28**, 3030–3043 (2017).
26. L. Schoneveld, A. Othmani, and H. Abdelkawy, “Leveraging recent advances in deep learning for audio-visual emotion recognition,” Pattern Recognit. Lett. **146**, 1–7 (2021).
27. J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” in *ICML,* (2011).
28. J. Summaira, X. Li, A. M. Shoib, S. Li, and J. Abdul, “Recent advances and trends in multimodal deep learning: A review,” arXiv preprint arXiv:2105.11087 (2021).
29. K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, “A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets,” The Vis. Comput. pp. 1–32 (2021).
30. T. Doherty, S. McKeever, N. Al-Attar, T. Murphy, C. Aura, A. Rahman, A. O’Neill, S. P. Finn, E. Kay, W. M. Gallagher *et al.*, “Feature fusion of raman chemical imaging and digital histopathology using machine learning for prostate cancer detection,” Analyst **146**, 4195–4211 (2021).
31. L. P. Rangaraju, G. Kunapuli, D. Every, O. D. Ayala, P. Ganapathy, and A. Mahadevan-Jansen, “Classification of burn injury using raman spectroscopy and optical coherence tomography: An ex-vivo study on porcine skin,” Burns **45**, 659–670 (2019).
32. L. Zhou, R. Griffith, and B. Gaeta, “Combining spatial and chemical information for clustering pharmacophores,” BMC bioinformatics **15**, 1–12 (2014).
33. A. Kumar and K. Y. Zhang, “Advances in the development of shape similarity methods and their application in drug discovery,” Front. chemistry **6**, 315 (2018).
34. S. Samaras, E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magoulianitis, A. Lalas, A. Dimou,

D. Zarpalas, K. Votis *et al.*, “Deep learning on multi sensor data for counter uav applications—a systematic review,” Sensors **19**, 4837 (2019).
35. L. Gómez-Chova, D. Tuia, G. Moser, and G. Camps-Valls, “Multimodal classification of remote sensing images: A review and future directions,” Proc. IEEE **103**, 1560–1584 (2015).
36. K. L. Law and R. C. Thompson, “Microplastics in the seas,” Science **345**, 144–145 (2014).
37. T. Takahashi, S. Yoshino, Y. Takaya, T. Nozaki, K. Ohki, T. Ohki, T. Sakka, and B. Thornton, “Quantitative in situ mapping of elements in deep-sea hydrothermal vents using laser-induced breakdown spectroscopy and multivariate analysis,” Deep. Sea Res. Part I Ocean. Res. Pap. p. 103232 (2020).
38. T. Fukuba, Y. Aoki, N. Fukuzawa, T. Yamamoto, M. Kyo, and T. Fujii, “A microfluidic in situ analyzer for ATP quantification in ocean environments,” Lab on a Chip **11**, 3508–3515 (2011).
39. R. G. de Vega, S. Goyen, T. E. Lockwood, P. A. Doble, E. F. Camp, and D. Clases, “Characterisation of microplastics and unicellular algae in seawater by targeting carbon via single particle and single cell icp-ms,” Anal. Chimica Acta **1174**, 338737 (2021).
40. C. B. Field, M. J. Behrenfeld, J. T. Randerson, and P. Falkowski, “Primary production of the biosphere: integrating terrestrial and oceanic components,” science **281**, 237–240 (1998).
41. P. G. Falkowski, “The role of phytoplankton photosynthesis in global biogeochemical cycles,” Photosynth. research **39**, 235–258 (1994).
42. L. A. Levin, K. Mengerink, K. M. Gjerde, A. A. Rowden, C. L. Van Dover, M. R. Clark, E. Ramirez-Llodra, B. Currie, C. R. Smith, K. N. Sato *et al.*, “Defining “serious harm” to the marine environment in the context of deep-seabed mining,” Mar. Policy **74**, 245–259 (2016).
43. A. C. Vivekanand, S. Mohapatra, and V. K. Tyagi, “Microplastics in aquatic environment: Challenges and perspectives,” Chemosphere **282**, 131151 (2021).
44. N. M. Burns and J. Watson, “Robust particle outline extraction and its application to digital in-line holograms of marine organisms,” Opt. Eng. **53**, 112212 (2014).
45. Z. Liu, T. Thevar, T. Takahashi, N. Burns, T. Yamada, M. Sangekar, D. Lindsay, J. Watson, and B. Thornton, “Unsupervised feature learning and clustering of particles imaged in raw holograms using an autoencoder,” J. Opt. Soc. Am. A **38**, 1570–1580 (2021).
46. N. Akhter, G. Min, J. W. Kim, and B. H. Lee, “A comparative study of reconstruction algorithms in digital holography,” Optik **124**, 2955–2958 (2013).
47. T. Latychevskaia and H.-W. Fink, “Practical algorithms for simulation and reconstruction of digital in-line holograms,” Appl. optics **54**, 2424–2434 (2015).
48. B. Efron, “Bootstrap methods: another look at the jackknife,” in *Breakthroughs in statistics,* (Springer, 1992), pp. 569–593.
49. X. Zhang, Z. Du, R. Zheng, Z. Luan, F. Qi, K. Cheng, B. Wang, W. Ye, X. Liu, C. Lian, C. Chen, J. Guo, Y. Li, and J. Yan, “Development of a new deep-sea hybrid Raman insertion probe and its application to the geochemistry of hydrothermal vent and cold seep fluid,” Deep. Res. Part I: Oceanogr. Res. Pap. **123**, 1–12 (2017).
50. V. Mazet, “Background correction,” MATLAB Central File Exchange. Retrieved March 7, 2022 (2022). Https://www.mathworks.com/matlabcentral/fileexchange/27429-background-correction.
51. V. Nava, M. L. Frezzotti, and B. Leoni, “Raman spectroscopy for the analysis of microplastics in aquatic systems,” Appl. Spectrosc. **75**, 1341–1357 (2021).
52. R. Withnall, B. Z. Chowdhry, J. Silver, H. G. Edwards, and L. F. de Oliveira, “Raman spectra of carotenoids in natural products,” Spectrochimica acta part a: molecular biomolecular spectroscopy **59**, 2207–2212 (2003).
53. S. Roberts and J. Murray, “Characterization of cement mineralogy in agglutinated foraminifera (protista) by raman spectroscopy,” J. Geol. Soc. **152**, 7–9 (1995).
54. G. Dong, G. Liao, H. Liu, and G. Kuang, “A review of the autoencoder and its variants: A comparative perspective from target recognition in synthetic-aperture radar images,” IEEE Geosci. Remote. Sens. Mag. **6**, 44–68 (2018).
55. T. Zeng, Y. Zhu, and E. Y. Lam, “Deep learning for digital holography: a review,” Opt. Express **29**, 40572–40593 (2021).
56. D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders,” arXiv preprint arXiv:2003.05991 (2020).
57. B. Melit Devassy, S. George, and P. Nussbaum, “Unsupervised clustering of hyperspectral paper data using t-sne,” J. Imaging **6**, 29 (2020).
58. K. Adem, “Diagnosis of breast cancer with stacked autoencoder and subspace knn,” Phys. A: Stat. Mech. its Appl. **551**, 124591 (2020).
59. T. Yamada, A. Prügel-Bennett, and B. Thornton, “Learning features from georeferenced seafloor imagery with location guided autoencoders,” J. Field Robotics **38**, 52–67 (2021).
60. T. Yamada, M. Massot Campos, A. Prugel-Bennett, O. Pizarro, S. B. Williams, and B. Thornton, “Guiding labelling effort for efficient learning with georeferenced images,” IEEE Transactions on Pattern Analysis Mach. Intell. (2022).
61. Z. Wang and Y. Wang, “Extracting a biologically latent space of lung cancer epigenetics with variational autoencoders,” BMC bioinformatics **20**, 1–7 (2019).
62. J. Opitz and S. Burst, “Macro F1 and macro F1,” arXiv preprint arXiv:1911.03347 (2019).