APL Photonics

Microplastic pollution assessment with digital holography and zero-shot learning

Cite as: APL Photonics **7**, 076102 (2022); https://doi.org/10.1063/5.0093439
Submitted: 28 March 2022 • Accepted: 13 June 2022 • Accepted Manuscript Online: 14 June 2022 • Published Online: 01 July 2022

Yanmin Zhu, Hau Kwan Abby Lo, Chok Hang Yeung, et al.

ARTICLES YOU MAY BE INTERESTED IN

Few-shot transfer learning for holographic image reconstruction using a recurrent neural network
APL Photonics **7**, 070801 (2022); https://doi.org/10.1063/5.0090582

Low-loss fiber grating coupler on thin film lithium niobate platform
APL Photonics **7**, 076103 (2022); https://doi.org/10.1063/5.0093033

Optical parametric wideband frequency modulation
APL Photonics **7**, 066106 (2022); https://doi.org/10.1063/5.0092969

AIP Publishing

Microplastic pollution assessment with digital holography and zero-shot learning

Cite as: APL Photon. 7, 076102 (2022); doi: 10.1063/5.0093439
Submitted: 28 March 2022 • Accepted: 13 June 2022 •
Published Online: 1 July 2022

Yanmin Zhu,[1] Hau Kwan Abby Lo,[2] Chok Hang Yeung,[2,3] and Edmund Y. Lam[1,a)]

AFFILIATIONS

[1] Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[2] Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[3] Centre for Water Technology and Policy, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

[a)] Author to whom correspondence should be addressed: elam@eee.hku.hk

ABSTRACT

Microplastic (MP) pollution poses severe environmental problems. Developing effective imaging tools for the identification and analysis of MPs is a critical step to curtail their proliferation. Digital holographic imaging can record the morphological and refractive index information of such small plastic fragments, yet due to the heterogeneous sampling environments and variations in the MP shapes, traditional supervised learning methods are of limited use. In this work, we pioneer a zero-shot learning method that combines the holographic images with their semantic attributes to identify the MPs in heterogeneous samples, even if they have not appeared in the training dataset. It makes use of the attention mechanism for image feature extraction and the Kullback–Leibler divergence both to alleviate the domain shift problem and to guide the training of the mapping function. Experimental results demonstrate the effectiveness of our approach and the potential use in a wide variety of environmental pollution assessments.

I. INTRODUCTION

Microplastic (MP) particles refer to plastic fragments with a diameter less than 5 mm.[1] They are increasingly recognized as severe sources of pollution. Many domestic and industrial activities, such as the shedding of fine fibers during laundry, marine aquaculture, and the use of daily necessities, such as toothpaste and facial cleansers, which contain plastic microbeads, produce a large amount of MPs.[2] In addition, plastic products, such as disposable lunch boxes and bags, may also break down into MPs due to incomplete recycling, degradation,[3] and physical abrasion.[4] Since the COVID-19 pandemic, such pollution has been exacerbated by the massive production of plastic medical wastes, such as disposable syringes and face masks. To make things worse, MPs may be ingested by marine organisms and further endanger human health through their spread in the food chain.[2] Therefore, it is imperative to control the pollution due to MPs.

A critical step is to identify the type and abundance of MPs in environmental samples, where they may be mixed with dust, dirt, and other small objects. In the laboratories, a common approach is to pre-filter the samples and pass them through imaging devices, such as Fourier-transform infrared spectroscopy (FT-IR),[5] Raman spectroscopy,[6] or even scanning electron microscopy (SEM).[7] More recently, digital holography (DH) is also emerging as a viable alternative.[6,8–14] It has several advantages, such as being a non-contact and high-throughput imaging method with relatively loose requirements for the imaging environment, and it does not require complicated pre-processing and preparation of samples. The possibility of using it in the field and coupling with quick image analysis sets it apart from other lab-based methods.[6,10,15,16] This is because the spatial interference between the reference light and the object light forms distinct coherence fringes and holograms, thereby capturing the morphological and refractive index information of the micro-objects, and computational processing can then lead to effective MP identification.[10]

In recent years, deep learning-based DH identification and classification methods using supervised training involving well-labeled datasets have made great progress in a variety of applications.[17–19] In these datasets, each image has a corresponding category label so that the classification network, such as a convolutional neural network (CNN)[20,21] or VGGNets,[9,22,23] learns the relationship between the image features and the labels, and matches the new input images in the test datasets based on such relationships. The classification performance of the network, to a large extent, depends on the number of image–label pairs and whether the images are correctly labeled. However, in many situations, the laborious task of image labeling, especially in some biomedical[19,20,24] and environmental[17,25–27] applications, needs to be conducted by professional staff and is very time-consuming. In addition, some objects, such as MPs, have distinct feature differences even though they are in the same class,[4] making it difficult for networks to learn the mapping functions and classify the unseen input images depending only on image–label pairs.

Zero-shot learning (ZSL) is an emerging approach that can classify unseen or unknown object classes with the help of auxiliary information, such as attributes, text, and class typology information.[28,29] In ZSL, the network learns the relationship between the data and their labels by training and, subsequently, can match the test data with unseen class labels accurately. As such, the classification performance is not solely reliant on the class labels but is enhanced by the help of some auxiliary information to reduce the need for complicated manual labeling.

In this paper, we propose a ZSL method to recognize the MPs in heterogeneous samples based only on their digital holographic images and semantic attributes. We build a semantic space to assist the classification and reduce the error caused by the MP physical feature variations. In addition, we make use of the attention mechanism to mimic the cognitive capabilities of humans in order to focus on the object features instead of the background area. Furthermore, to alleviate the domain shift problem, which is common in ZSL, we introduce the Kullback–Leibler divergence to guide the learning of the mapping function. The effectiveness of the proposed method is demonstrated both on the DH dataset and the extended digital image MP dataset. Experimental results show a substantial performance improvement in MP recognition.

The rest of this paper is organized as follows. In Sec. II, we outline the current progress in MP identification. We also introduce the related work of learning-based digital holographic imaging for MP classification, ZSL and the semantic space, and the attention mechanism. In Sec. III, we pinpoint the present challenges and provide a detailed description and explanation of our method. In Sec. IV, we introduce the experimental setup and our main results, together with ablation experiments to validate the effectiveness of our method. Finally, in Sec. V, we give a brief summary and discuss the future directions following this work.

II. RELATED WORK

A. Learning-based microplastic particles classification with digital holography

In holographic imaging, the fringe patterns indicate the optical path difference from the samples. This characteristic allows us to compute the refractive index and retrieve texture information, making it possible to identify much information about the samples with deep learning. Consequently, several groups have explored the use of learning-based holographic imaging to detect and identify MP particles. Bianco *et al.*[11] demonstrate the feasibility of using machine learning to classify plastic fragments. Their method consists of an artificial holographic feature extraction and an image classification using the support vector machine. However, a limitation of this technique lies in the challenge of extracting effective features for image classification.

More recently, deep learning has been applied to identify and quantify the MPs among different mixtures and in challenging imaging scenarios.[9,12,30,31] Although the effectiveness of learning-based classification has been demonstrated on electron microscopic MP images,[30,32] FT-IR spectra,[5,33] vibration spectra,[34] and dark field microscopic images,[35] the application of learning-based methods in holographic MP imaging is not as common, despite various advantages, such as high resolution, high speed, and non-contact. One primary reason has been that a well-labeled holographic image dataset requires costly manpower and material resources to build. Zhu *et al.* partially overcame this limitation and constructed a dataset of holographic MP images and put it in open source[36] to achieve real-time MP quantification[9] and detection.[12] In addition, they also develop methods with transfer learning[25] and generative adversarial networks[25] to alleviate the misclassification caused by the imbalanced distribution of image classes in the MP holographic dataset.

B. Zero-shot learning and the semantic space

ZSL is developed to classify the unseen classes in the target based on the seen classes in the source by generalizing the image features from the latter to the former.[37] As shown in Fig. 1, a mapping function relates the visual space to the semantic space. The expert-designed attributes provide reliable descriptions for the image classes, improving classification accuracy effectively. Because the mapping is trained on the seen classes and then applied to the unseen classes, the key issue in ZSL is figuring out how to learn one such function that is both general and accurate.

Several models are commonly used to learn the mapping functions, including both the direct attribute prediction (DAP)[28] and the indirect attribute prediction (IAP).[38] However, since there is no intersection between the seen classes in the source and the unseen classes in the target, the mapping function may suffer from serious domain shifting when learning only on the source data.[39]

C. Attention mechanism for image feature extraction

The conventional CNNs uniformly extract image features from all regions for further classification. Such an approach is computationally costly and inefficient. Instead, the attention mechanism is designed to mimic the visual attention distribution of human beings when they are viewing an image. Humans automatically focus more attention on attractive areas than on the whole image.[40] Similarly, with the attention mechanism, the features extracted from various regions of the image are assigned different weights according to a weight mask. The effectiveness of attention-based feature extraction and image processing has been demonstrated in several recent works.[23,26,41]

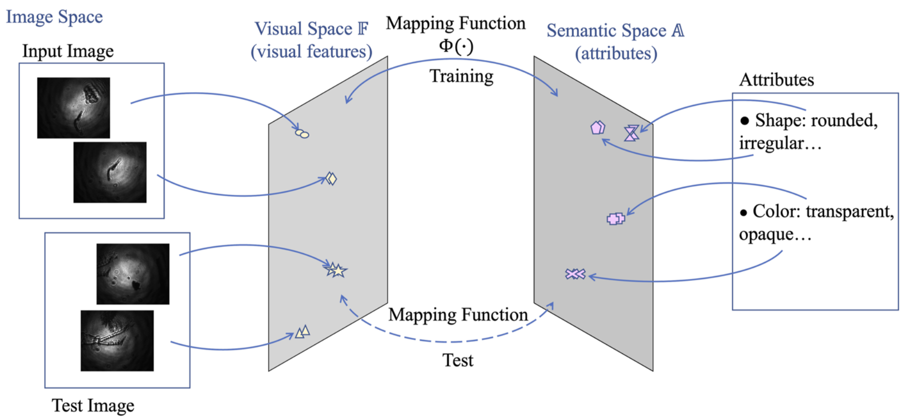

FIG. 1. Schematic illustration of the feature projection between the visual space $\mathbb{F}$ and the semantic space $\mathbb{A}$. The input images are represented by the visual features in the former. The attributes are transferred to it by the mapping function $\Phi(\cdot)$, which is learned during the training process.

III. METHOD

A. Problem definition

Suppose we have data belonging to one of two sets: dataset $\mathscr{D}_s$, with N_s images from the seen classes, and dataset $\mathscr{D}_u$, with N_u images from the unseen classes.

In the former, assume that there are C_s seen classes. For the jth image, which we denote as $\mathbf{x}_j^s$, it has a class label c_j^s and a set of attributes $\mathscr{A}_j$. Note that the images in the same class are described by the same attributes. Thus, we have

$$\mathscr{D}_s = \left\{\mathbf{x}_j^s, c_j^s, \mathscr{A}_j\right\} \quad \text{for } j = 1, \ldots, N_s. \tag{1}$$

Similarly, in the latter, assume that there are C_u unseen classes. For the jth image, which we denote as $\mathbf{x}_j^u$, it has no label but there is still a set of attributes $\mathscr{A}_j$. Again, the images in the same class should have the same attributes. Correspondingly, we have

$$\mathscr{D}_u = \left\{\mathbf{x}_j^u, \mathscr{A}_j\right\} \quad \text{for } j = 1, \ldots, N_u. \tag{2}$$

The goal of ZSL is to predict, for image $\mathbf{x}_j^u$, the class $c_j^u \in C_u$ with the help of $\mathscr{A}_j$. Note that there is no overlap between the seen classes C_s and the unseen classes C_u, i.e., $C_s \cap C_u = \emptyset$. As illustrated in Fig. 1, we aim to achieve this by learning a mapping function $\Phi(\cdot)$ on the seen dataset $\mathscr{D}_s$ and aligning the feature centers in the semantic space $\mathbb{A}$ to the centers in the visual space $\mathbb{F}$.

B. Method

Attention-based feature extraction. The images in both datasets $\mathscr{D}_s$ and $\mathscr{D}_u$ can be represented by the extracted features in the visual space. Thus, we calculate the feature center of each class to represent its location. In addition, the attributes of each class are coded as vectors and represented by attribute centers in the semantic space.

To achieve the goal in Sec. III A, we apply and then adjust the attention mechanism to localize the interest region of the input images, where we keep the weights within the object and scale down those in the less dominant areas. Taking advantage of the superior capability of the convolutional block attention module (CBAM) on weight scaling,[41] we use ResNet-101[42] as the backbone network and augment it with CBAM in each of the convolution layers to focus on the more important image regions. The whole process is illustrated in Fig. 2(a).

More specifically, as shown in Figs. 2(b) and 2(c), the input feature map $\mathscr{F}_i$ from the previous layer is first calculated by the convolution layer (conv). Next, the generated feature map, labeled as $\mathscr{F}_g$, is refined by the CBAM. Rather than directly scaling the weights in the width, height, and channel dimensions, CBAM decomposes the 3D feature tensors and computes the 1D channel attention map ($\mathscr{M}_c$), denoted as

$$\mathscr{F}_c = \mathscr{M}_c \odot \mathscr{F}_g, \tag{3}$$

where we use $\odot$ to represent the element-wise multiplication. The intermediate feature map $\mathscr{F}_c$ is then multiplied by the 2D spatial attention map ($\mathscr{M}_s$), written as

$$\mathscr{F}_r = \mathscr{M}_s \odot \mathscr{F}_c, \tag{4}$$

where $\mathscr{F}_r$ is the refined feature map. Previous experimental results show that this multiplication mechanism can reduce replication in the computation.[41]

In the overall feature extraction network shown in Fig. 2(a), for each input image $\mathbf{x}_j \in \mathscr{D}_s \cup \mathscr{D}_u$, a d-dimensional feature vector $\mathbf{f}_j \in \mathbb{R}^{d\times 1}$ is extracted, and all input images are embedded into the visual space $\mathbb{F}$ in this form. For each class, a feature center $\mathbf{v}_c^{\mathrm{I}}$ (where $c \in C_s \cup C_u$) is defined as the mean value of all the image feature vectors in this class, where the superscript I refers to the image. Similarly, in the semantic space $\mathbb{A}$, a mapping function $\Phi(\cdot)$ is used for transferring the attribute $\mathscr{A}_j$ and generating the feature centers $\mathbf{v}_c^{\mathrm{A}}$, which is a trainable network containing two fully connected layers and nonlinear Leaky ReLU activation functions. The transfer process can be represented as

$$\mathbf{v}_c^{\mathrm{A}} = \phi(\mathscr{A}_j). \tag{5}$$

We denote the feature centers of the images as the visual center, and the attribute feature centers as the semantic center. As such, we can classify the images of the unseen classes based on the seen class images by ZSL, provided that the visual center aligns with the semantic center.

Feature center alignment. ZSL is susceptible to the domain shift problem,[39] which stems from the image feature distribution gaps between the seen and unseen classes. The mapping function is learned on the former and applied to the latter directly, which results

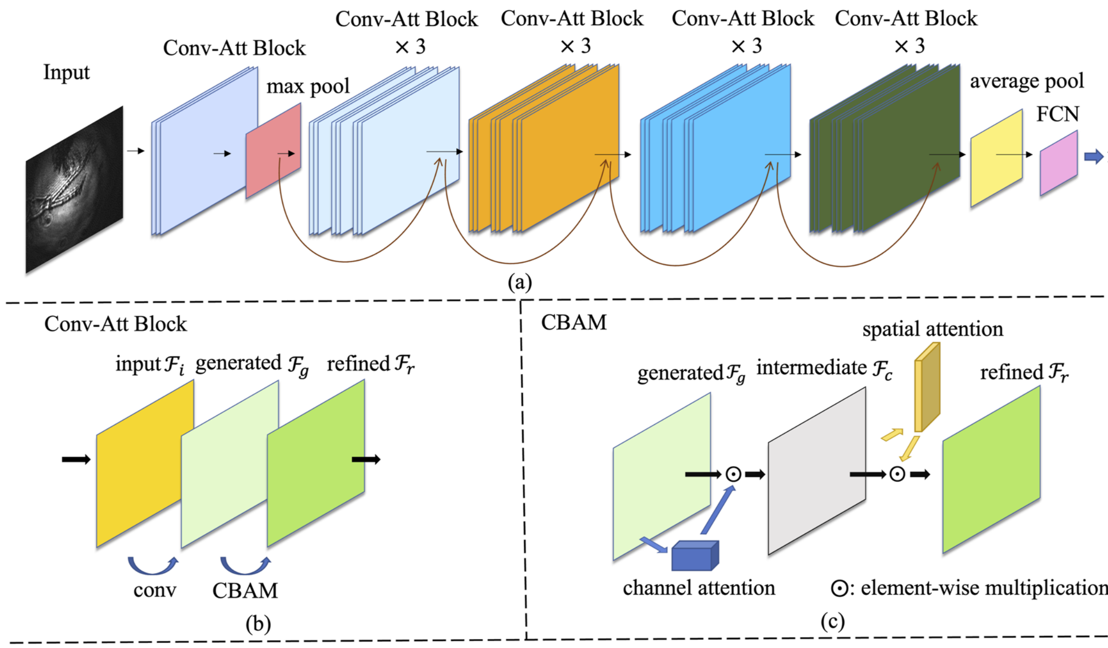

FIG. 2. The structure diagrams of (a) the feature extraction network, (b) the Conv-Att block, and (c) the convolutional block attention module (CBAM). The feature extraction network has ResNet-101 as the backbone and consists of a max-pooling layer, Conv-Att blocks, an average pooling layer, and a fully connected layer (FCN). The input image is processed by the feature extraction network, and a feature vector $\mathbf{f}_j$ is generated as the output.

in the loss of network generalization and a drop in performance. This is illustrated in Fig. 3. However, under ideal circumstances, the visual feature centers of the unseen classes should be perfectly aligned with the attribute centers by the mapping function. Making use of the work by Stacke *et al.*,[43] which quantifies the distribution difference to accomplish domain adaption, we employ the Kullback–Leibler (KL) divergence[44] as the loss function to learn a soft assignment mapping function from the attribute feature center set V_A to the image feature center set V_{I}. This is defined as

$$\mathcal{L}_{KL} = \mathcal{D}_{KL}(V_A \parallel V_{\mathrm{I}}) = \sum_{c=1}^{C_u} \mathbf{v}_c^{\mathrm{A}} \log \frac{\mathbf{v}_c^{\mathrm{A}}}{\mathbf{v}_c^{\mathrm{I}} + \epsilon}, \tag{6}$$

where ϵ is a small value to prevent the denominator from becoming zero. Along with the KL divergence loss, which computes the distribution difference, the mean square error (MSE) loss also minimizes the distance between these two kinds of centers. This is given by

$$\mathcal{L}_{MSE} = \frac{1}{C_s} \sum_{c=1}^{C_s} \|\mathbf{v}_c^{\mathrm{A}} - \mathbf{v}_c^{\mathrm{I}}\|_2^2. \tag{7}$$

The overall loss function is defined as

$$\mathcal{L} = \mathcal{L}_{MSE} + \lambda \mathcal{L}_{KL}, \tag{8}$$

where λ adjusts the relative weight of the two terms.

IV. EXPERIMENT

A. Experimental setup and dataset

We evaluate our method on its ability to identify microplastics among heterogeneous samples. To obtain the discriminative image features of MPs,[10,46] we use a specially designed single-shot off-axis digital holographic system[47] to record the images of these samples. This system has the ability to record holograms with different polarization states in a single shot, which can then be used to synthesize high-resolution holograms.[48] The physical image and the corresponding optical principle diagram of the system setup are both shown in Figs. 4(a) and 4(b), respectively. The light, which comes from the laser beam, is expanded by the beam expander (BE) and separated into the object light and reference light by several beam splitters (BSs). The former then goes through the samples and interferes with the latter. The resulting holograms are enlarged by the microscopic objective (MO) and recorded by an electronic sensor. We use a neutral density filter (NDF) for the light density control, and several polarizers to adjust the polarization states of the reference light. The holographic images are recorded with 1280 × 1280 pixels, resized to 224 × 224 pixels to be used as input to the network while the feature vectors are preserved during the training and testing stages.

Altogether, we collect in this way a dataset containing 1300 images with eight classes, namely, fragment, foam, film, fiber, fiber and fragment, sand, shells and bones, and others. These attribute descriptions and category definitions follow the illustrations by Rocha-Santos and Duarte,[4] Chap. 3.

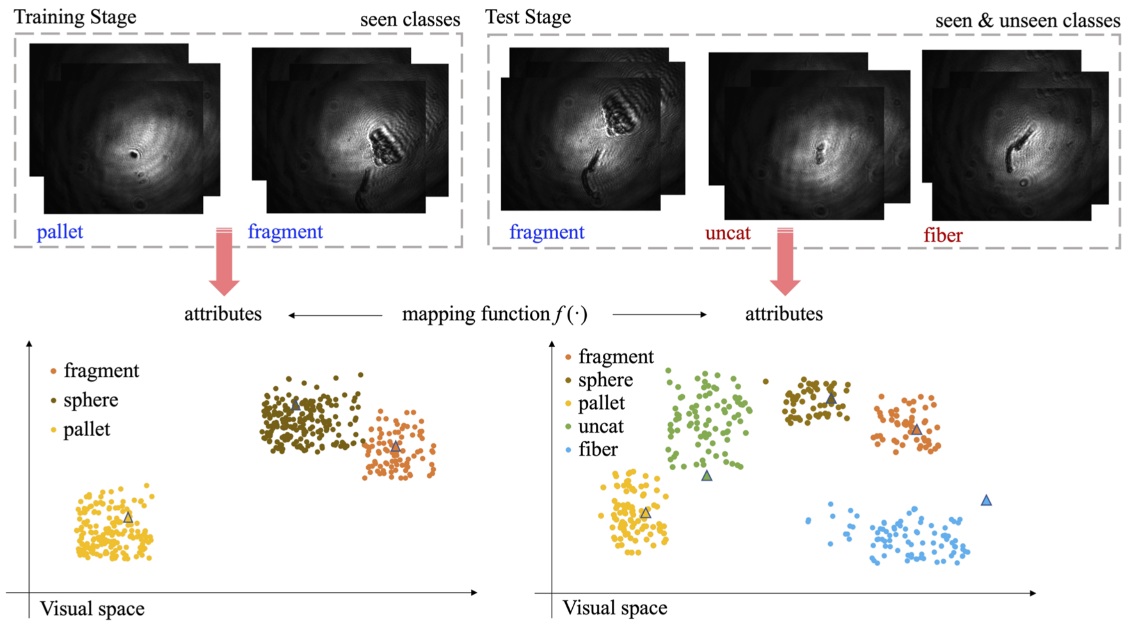

FIG. 3. The schematic illumination of domain shift in ZSL. Because of the disjoint nature and feature distribution differences of the seen and unseen classes, when the mapping function is learned on the former and directly applied on the latter, there can be a domain shift, and the feature centers (represented by the triangles) may stray from their real locations.

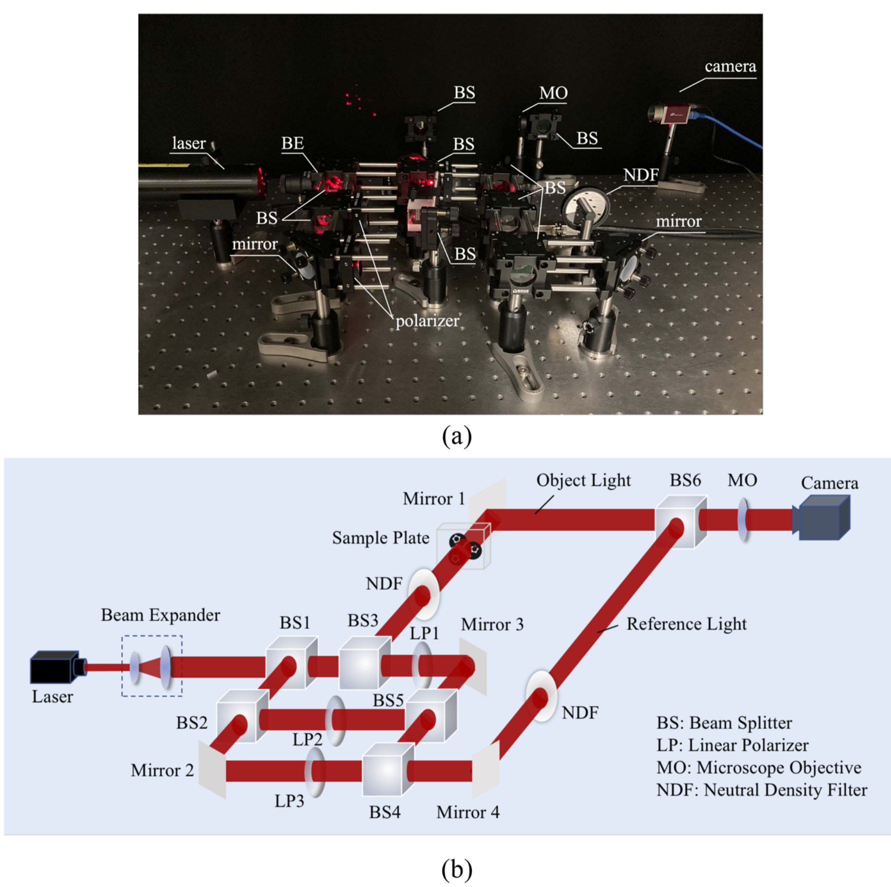

FIG. 4. The single-shot off-axis digital holographic system, which is used to record the image dataset. (a) The physical image of the experimental setup. BE: beam expander, BS: beam splitter, MO: microscope objective, and NDF: neutral density filter. (b) The optical schematic of the experimental setup.

B. Implementation details

The network is implemented in Python. The training process is running on an Intel Xeon CPU with 12 GB RAM and accelerated by a Tesla K80 GPU. The factor λ is fine-tuned in the range of $[10^{-3}, 10^{-4}]$ with cross-validation.

C. Qualitative analysis

Inspection of attention-based feature extraction. We plot the activation heat map in Fig. 5 to visualize the interest region of the feature extraction with the network. Focus regions show higher heat values in these plots.[49] It can be seen that the high heat value regions appear around the microparticles, which indicates the effectiveness of the attention-based feature extraction mechanism in scaling the weights of the feature maps according to the image contents. By putting less focus on the background regions, our network has the capability to extract discriminative features, which are then evaluated in the MP classification tasks.

Feature center alignment. We randomly select some sample images from the holographic dataset and plot their feature distributions in order to visualize the capability of the feature center alignment of our method. As shown in Fig. 6, the semantic centers of each class, which are generated by the mapping function, are located closely to their visual centers. This result shows the effectiveness of the network training strategy with the KL loss function in learning the mapping function.

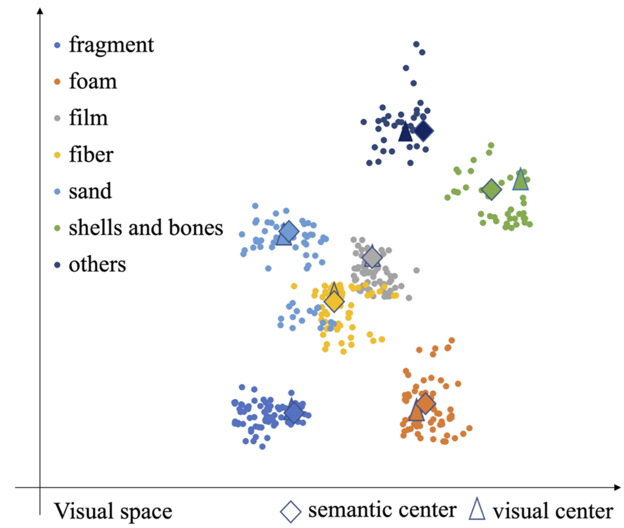

FIG. 6. Visualization of the feature distribution of randomly selected images from the dataset using t-SNE.[45] The semantic centers generated by the mapping function locate closely to the visual centers.

D. Experimental results and discussions

We measure the average per-class top-1 accuracy on both the seen and unseen classes to evaluate the classification performance, which is denoted by A_s and A_u, respectively. Because the accuracies of the seen classes are much more prominent than those of the unseen classes in ZSL,[37] we calculate the harmonic mean accuracy to soften the impact of the imbalance between the two, defined as

$$H = \frac{2 \cdot A_s \cdot A_u}{A_u + A_s}. \tag{9}$$

TABLE I. Experimental results of the ablation study for attention-based feature extraction mechanism. A_s and A_u represent the top-1 accuracy on the seen and unseen classes, respectively. H is their harmonic mean. The results are the average values based on ten repeated experiments.

Methods	A_s (%)	A_u (%)	H (%)
ResNet-101	74.02	52.37	61.34
Our method	88.04	71.61	78.98

We first evaluate the effectiveness of the attention based feature extraction mechanism. We compare the classification performance of our method and the ResNet-101 network, which has the same backbone model but without the attention mechanism, on the heterogeneous MP dataset. As reported in Table I, the experimental

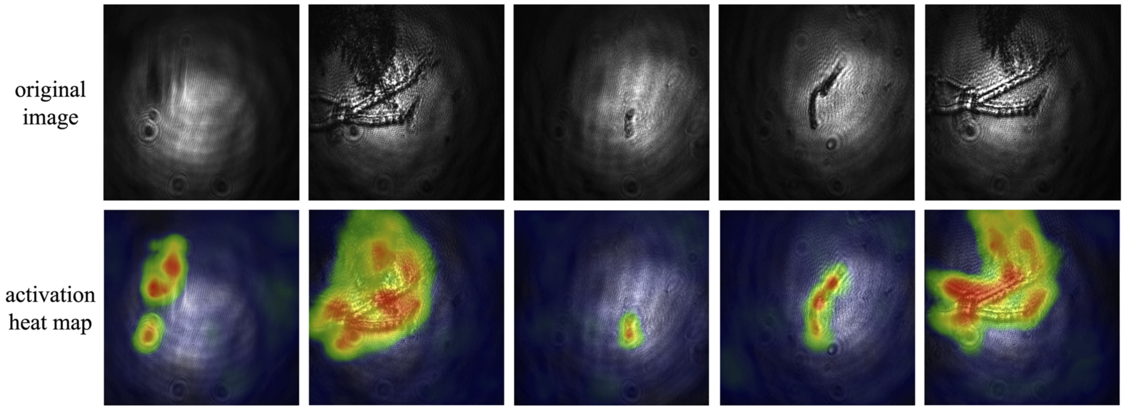

FIG. 5. Visualization of the activation heat maps of randomly selected images from the dataset. The proposed attention-based feature extraction mechanism scales up the weights of the interest region in the image.

TABLE II. Experimental results in top-1 accuracy and the harmonic mean accuracy with different distance loss functions. A_s and A_u represent the top-1 accuracy on the seen and unseen classes, respectively. H is their harmonic mean. The notations $\mathscr{L}_{cf}$ and $\mathscr{L}_W$ refer to the Chamfer distance and the Wasserstein distance, respectively. The results are the average values based on ten repeated experiments.

Loss function	A_s (%)	A_u (%)	H (%)
$\mathscr{L}_{MSE}$	66.54	50.21	57.23
$\mathscr{L}_{MSE} + \mathscr{L}_{cf}$	73.57	55.74	63.43
$\mathscr{L}_{MSE} + \mathscr{L}_W$	76.35	59.63	66.96
Our method	87.84	70.97	78.51

results of our method on both the seen and unseen classes show significant improvements. The harmonic mean records a 28.76% gain compared to the ResNet-101 network, which demonstrates the significance of the attention mechanism on the feature extraction and the MP classification tasks.

As discussed earlier, ZSL generally suffers from the domain shift problem. As shown in Fig. 6, there is a variation between the visual center and the semantic center of each class, which may raise from the mismatch between the visual features and the semantic features. Here, we analyze the performance of our method in alleviating this issue by observing the experimental results with different loss functions for the MP classification tasks. The Chamfer distance[50] and the Wasserstein distance[51] are selected for the comparison experiments, as they are both commonly used distance measurements.[52] In addition, we also report the results with only the MSE loss as the control group. As given in Table II, our method achieves better results on both top-1 accuracies and their harmonic mean over the other three distance functions. Combining with the visualization results in Fig. 6, we can conclude that the KL divergence loss function makes a positive contribution in dealing with the domain shift problem and classifying heterogeneous MP samples.

In reality, MPs have high variance in their shape and morphology. No existing database can cover all the varieties. It, therefore, has crucial significance for researchers if the network can give correct

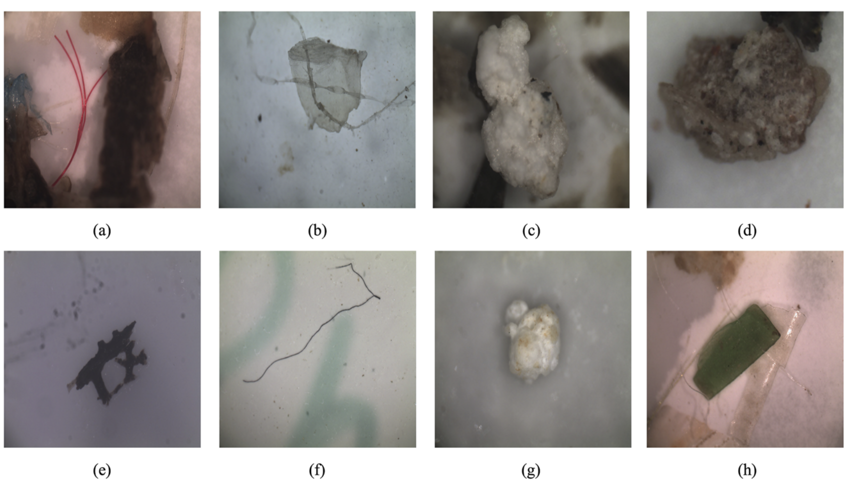

FIG. 7. Sample images of the seen classes from the heterogeneous microplastic digital image dataset. The classes of the sample images are (a) fiber, (b) film, (c) foam, (d) sand, (e) shells and bones, (f) fiber, (g) foam, and (h) fragment.

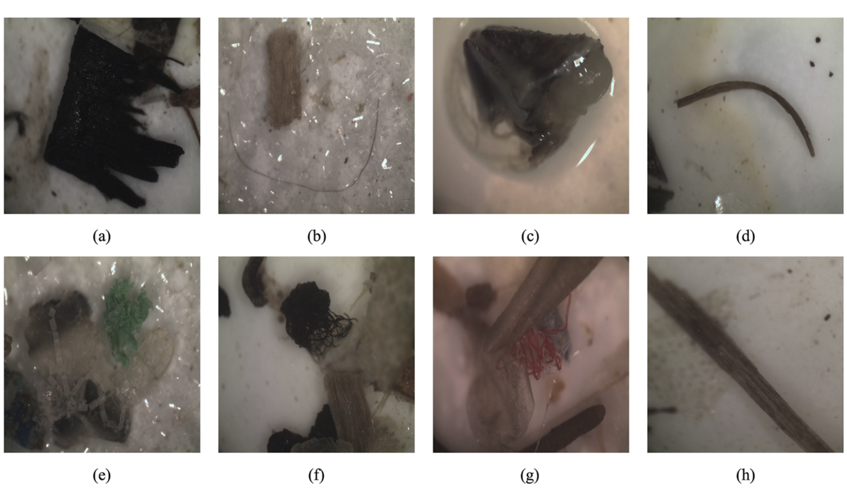

FIG. 8. Sample images of the unseen classes from the heterogeneous microplastic digital image dataset. The classes of the sample images are (a) coal dust and fly ash, (b) pellet, (c) micro-organisms, (d) fiber and fiber bundle, (e)–(g) fiber bundle, and (h) algae and plant matter.

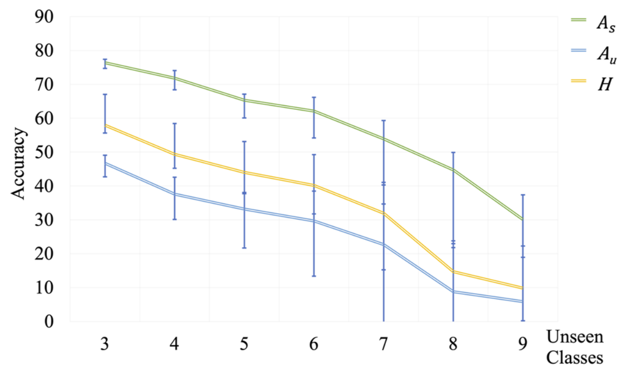

FIG. 9. Accuracy of the seen (A_s) and unseen (A_u) classes, and the harmonic mean (H) vs the number of unseen classes on the microplastic digital image dataset.

predictions for new MP samples. In addition, some datasets contain conventional images rather than holograms of the MPs, and it is highly desirable if the network can also deliver good results on such inputs. To this end, we conduct experiments on a digital MP image dataset collected by ourselves in multiple sampling sites in Hong Kong.

We split the images here into the same classes as the holographic image dataset by some professional researchers. As shown in Figs. 7 and 8, the digital image dataset has more complex content, and it is more difficult to make predictions. Natural particles, such as organic matters and inert minerals,[53] mix with the MPs, resulting in occlusion and confusion for the MP recognition. For this challenging task, we show the experimental results with the change in the number of unseen classes in Fig. 9. It is worth noting that our method achieves nearly 60% in the harmonic mean H on the digital image dataset. Moreover, it shows a slow drop with an increase in the number of unseen classes. These experimental results indicate a potential value of our method to classify and identify the heterogeneous MP samples in the field.

V. CONCLUSION

In this work, we propose a zero-shot learning method for the heterogeneous microplastic sample identification and classification task based on their digital holographic images. Different from the traditional supervised learning classification methods, ZSL delivers notable results on image classification even though the samples are never seen during the training process. By leveraging the attention mechanism and Kullback–Leibler divergence, our method shows a substantial performance gain in extracting the image feature and reducing the domain shift. The experiments are not only conducted on the holographic image dataset, which contains the morphological and refractive index information of the MPs, but also on the more general MP digital image dataset. This work is significant in assisting the understanding of the MP pollution and distribution. It also provides insights into the potential capability of ZSL to identify MPs in the field intelligently.

ACKNOWLEDGMENTS

The work was supported, in part, by the Research Grants Council of Hong Kong (Grant Nos. GRF 17201620 and RIF R7003-21) and the Environment and Conservation Fund (ECF Project No. 109/2019).

We wish to acknowledge the support of Dr. Tianjiao Zeng for helpful comments on this work.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yanmin Zhu: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Hau Kwan Abby Lo:** Data curation (supporting); Writing – review & editing (supporting). **Chok Hang Yeung:** Data curation (supporting); Funding acquisition (lead); Investigation (lead); Supervision (supporting); Writing – review & editing (supporting). **Edmund Y. Lam:** Conceptualization (supporting); Funding acquisition (lead); Investigation (lead); Methodology (supporting); Resources (supporting); Supervision (lead); Writing – review & editing (supporting).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] K. L. Law and R. C. Thompson, "Microplastics in the seas," Science **345**, 144–145 (2014).

[2] A. D. Vethaak and J. Legler, "Microplastics and human health," Science **371**, 672–674 (2021).

[3] S. Allen, D. Allen, V. R. Phoenix, G. Le Roux, P. Durántez Jiménez, A. Simonneau, S. Binet, and D. Galop, "Atmospheric transport and deposition of microplastics in a remote mountain catchment," Nat. Geosci. **12**, 339–344 (2019).

[4] T. A. Rocha-Santos and A. C. Duarte, *Characterization and Analysis of Microplastics* (Elsevier, 2017), Chap. 3.

[5] V. H. da Silva, F. Murphy, J. M. Amigo, C. Stedmon, and J. Strand, "Classification and quantification of microplastics (<100 μm) using a focal plane array–Fourier transform infrared imaging system and machine learning," Anal. Chem. **92**, 13724–13733 (2020).

[6] 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).

[7] D. Elkhatib and V. Oyanedel-Craver, "A critical review of extraction and identification methods of microplastics in wastewater and drinking water," Environ. Sci. Technol. **54**, 7037–7049 (2020).

[8] Y. Zhu, C. Hang Yeung, and E. Y. Lam, "Digital holographic imaging and classification of microplastics using deep transfer learning," Appl. Opt. **60**, A38–A47 (2021).

[9]Y. Zhu, C. Hang Yeung, and E. Y. Lam, "Microplastic pollution monitoring with holographic classification and deep learning," J. Phys.: Photonics **3**, 024013 (2021).
[10]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).
[11]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).
[12]Y. Zhu, C. H. Yeung, and E. Y. Lam, "Automatic detection of microplastics by deep learning enabled digital holography," in *Digital Holography and Three-Dimensional Imaging* (Optical Society of America, 2020), p. HTu5B.1.
[13]V. Bianco, P. Memmolo, P. Carcagnì, F. Merola, M. Paturzo, C. Distante, and P. Ferraro, "Probing micro-plastic items with coherent light enables their identification through digital holography," in *Digital Holography and Three-Dimensional Imaging* (Optical Society of America, 2019), pp. Th2A–2.
[14]Á. Barroso, B. Kemper, S. Ketelhut, S. Graß, J. Reiber, and J. Schnekenburger, "Multimodal optical detection and toxicity testing of microplastics in the environment," in *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XVII* (SPIE, 2019), Vol. 10881, pp. 204–209.
[15]Y. Zhu, C. H. Yeung, and E. Y. Lam, "Holographic classifier: Deep learning in digital holography for automatic micro-objects classification," in *2020 IEEE 18th International Conference on Industrial Informatics* (IEEE, 2020), Vol. 1, pp. 515–520.
[16]T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: A review," Opt. Express **29**, 40572–40593 (2021).
[17]Y. Zhang, Y. Lu, H. Wang, P. Chen, and R. Liang, "Automatic classification of marine plankton with digital holography using convolutional neural network," Opt. Laser Technol. **139**, 106979 (2021).
[18]Z. Ren, H. K.-H. So, and E. Y. Lam, "Fringe pattern improvement and super-resolution using deep learning in digital holography," IEEE Trans. Ind. Inf. **15**, 6179–6186 (2019).
[19]T. O'Connor, A. Anand, B. Andemariam, and B. Javidi, "Deep learning-based cell identification and disease diagnosis using spatio-temporal cellular dynamics in compact digital holographic microscopy," Biomed. Opt. Express **11**, 4491–4508 (2020).
[20]N. Meng, E. Y. Lam, K. K. Tsia, and H. K. So, "Large-scale multi-class image-based cell classification with deep learning," IEEE J. Biomed. Health Inf. **23**, 2091–2098 (2018).
[21]X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," IEEE Trans. Geosci. Remote Sens. **56**, 937–949 (2017).
[22]K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
[23]C. Sitaula and M. B. Hossain, "Attention-based VGG-16 model for COVID-19 chest X-ray image classification," Appl. Intell. **51**, 2850–2863 (2021).
[24]Z. Ren, E. Y. Lam, and J. Zhao, "Learning-based cell detection in digital pathology," in *2021 Conference on Lasers and ElectroOptics (CLEO)* (IEEE, 2021), pp. 1–2.
[25]Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holography with deep learning and generative adversarial networks for automatic microplastics classification," in *Holography, Diffractive Optics, and Applications* (SPIE, 2020), Vol. 11551, p. 115510A.
[26]P. Guan and E. Y. Lam, "Multistage dual-attention guided fusion network for hyperspectral pansharpening," IEEE Trans. Geosci. Remote Sens. **60**, 5515214 (2022).
[27]Z. Ren, E. Y. Lam, and J. Zhao, "Real-time target detection in visual sensing environments using deep transfer learning and improved anchor box generation," IEEE Access **8**, 193512–193522 (2020).
[28]C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 951–958.
[29]Z. Wan, D. Chen, and J. Liao, "Visual structure constraint for transductive zero-shot learning in the wild," Int. J. Comput. Vision **129**, 1893–1909 (2021).
[30]J. Lorenzo-Navarro, M. Castrillón-Santana, E. Sánchez-Nielsen, B. Zarco, A. Herrera, I. Martínez, and M. Gómez, "Deep learning approach for automatic microplastics counting and classification," Sci. Total Environ. **765**, 142728 (2021).
[31]B. Shi, M. Patel, D. Yu, J. Yan, Z. Li, D. Petriw, T. Pruyn, K. Smyth, E. Passeport, R. J. D. Miller, and J. Y. Howe, "Automatic quantification and classification of microplastics in scanning electron micrographs via deep learning," Sci. Total Environ. **825**, 153903 (2022).
[32]C. Massarelli, C. Campanale, and V. F. Uricchio, "A handy open-source application based on computer vision and machine learning algorithms to count and classify microplastics," Water **13**, 2104 (2021).
[33]M. Kedzierski, M. Falcou-Préfol, M. E. Kerros, M. Henry, M. L. Pedrotti, and S. Bruzaud, "A machine learning algorithm for high throughput identification of FTIR spectra: Application on microplastics collected in the mediterranean sea," Chemosphere **234**, 242–251 (2019).
[34]H. d. M. Back, E. C. Vargas Junior, O. E. Alarcon, and D. Pottmaier, "Training and evaluating machine learning algorithms for ocean microplastics classification through vibrational spectroscopy," Chemosphere **287**, 131903 (2022).
[35]I. Ishmukhametov, L. Nigamatzyanova, G. Fakhrullina, and R. Fakhrullin, "Label-free identification of microplastics in human cells: Dark-field microscopy and deep learning study," Anal. Bioanal. Chem. **414**, 1297–1312 (2022).
[36]Y. Zhu, C. H. Yeung, and E. Y. Lam, "Holographic image dataset for microplastics," https://github.com/ymzhu19eee/dataset_microplastics, 2020.
[37]Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly," IEEE Trans. Pattern Anal. Mach. Intell. **41**, 2251–2265 (2018).
[38]C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," IEEE Trans. Pattern Anal. Mach. Intell. **36**, 453–465 (2013).
[39]Y. Fu, T. M. Hospedales, T. Xiang, and S. Gong, "Transductive multi-view zero-shot learning," IEEE Trans. Pattern Anal. Mach. Intell. **37**, 2332–2345 (2015).
[40]A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems* (Neural Information Processing Systems, 2017), pp. 5998–6008.
[41]S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "*CBAM: Convolutional block attention module*," in *Proceedings of the European Conference on Computer Vision* (Springer, 2018), pp. 3–19.
[42]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 770–778.
[43]K. Stacke, G. Eilertsen, J. Unger, and C. Lundström, "Measuring domain shift for deep learning in histopathology," IEEE J. Biomed. Health Inf. **25**, 325–336 (2020).
[44]F. Pérez-Cruz, "Kullback-Leibler divergence estimation of continuous distributions," in *IEEE International Symposium on Information Theory* (IEEE, 2008), pp. 1666–1670.
[45]L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," J. Mach. Learn. Res. **9**, 2579–2605 (2008).
[46]Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holographic microplastics detection and characterization in heterogeneous samples via deep learning," in *12th International Conference on Information Optics and Photonics* (SPIE, 2021), Vol. 12057, pp. 789–794.
[47]Y. Z. Zhu, T. Zeng, K. Liu, Z. Ren, and E. Y. Lam, "Full scene underwater imaging with polarization and untrained network," Opt. Express **29**, 41865–41881 (2021).
[48]C. Yuan, G. Situ, G. Pedrini, J. Ma, and W. Osten, "Resolution improvement in digital holography by angular and polarization multiplexing," Appl. Opt. **50**, B6–B11 (2011).
[49]A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional

networks," in *2018 IEEE Winter Conference on Applications of Computer Vision* (IEEE, 2018), pp. 839–847.

[50] M. A. Butt and P. Maragos, "Optimum design of chamfer distance transforms," IEEE Trans. Image Process. **7**, 1477–1484 (1998).

[51] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio, "Learning with a Wasserstein loss," in Proceedings of the 28th International Conference on Neural Information Processing Systems (2015), Vol. 2, pp. 2053–2061.

[52] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 605–613.

[53] A. A. Koelmans, P. E. Redondo-Hasselerharm, N. H. M. Nor, V. N. de Ruijter, S. M. Mintenig, and M. Kooi, "Risk assessment of microplastic particles," Nat. Rev. Mater. **7**, 138–152 (2022).