

## RESEARCH ARTICLE

# Deep Classification of Microplastics Through Image Fusion Techniques

PAOLO RUSSO<sup>1</sup> AND FABIANA DI CIACCIO<sup>2</sup><sup>1</sup>Department of Computer, Control and Management Engineering "Antonio Ruberti", Sapienza University of Rome, 00185 Rome, Italy<sup>2</sup>Department of Civil and Environmental Engineering, University of Florence, 50121 Florence, Italy


Corresponding author: Paolo Russo (paolo.russo@diag.uniroma1.it)

**ABSTRACT** Microplastics derived from fiber shredding are recognized by the scientific community as one of the main sources of microplastic water pollution, thus actualizing the need for techniques able to identify them with high accuracy. The recently released Holography Micro-Plastic Dataset offers the opportunity to test deep neural networks in their ability to distinguish between microplastics and other debris on a standard benchmark. The promising results obtained from the initial batch of experiments can be further improved through a combined approach which involves different image mapping techniques and recent state-of-the-art deep models. Within this framework, we analyze various image fusion schemas to merge the paired dataset images (amplitude and phase) into a single three-channel picture. We demonstrate that our proposed approach yields increased accuracy compared to both single-image data processing and other fusion techniques. Finally, the performance of our method is further enhanced by employing the Vision Transformer model as backbone, highlighting the effectiveness of the proposed approach in microplastics classification.

**INDEX TERMS** Deep learning, digital holography, image fusion, microplastics.

## I. INTRODUCTION

Plastic products are nowadays characterized by low costs, high strength/weight ratio, resistance to chemicals, temperature and light, and ease of manufacture [1]. These properties make them the preferred solution in a wide spectrum of applications: unsurprisingly, the 2 million tonnes per year of plastics produced in 1950 increased nearly 230-fold, reaching 460 million tonnes in 2019 [2]. The durability of these synthetic polymers makes them also highly resistant to degradation, causing them to be one of the main pollution sources in the environment [3]. In fact, even if recycling can be considered as one of the most effective solutions to this issue, all the plastic waste that is improperly managed becomes a significant risk due to dispersion into the environment. In order to give an idea of the magnitude of this problem, in 2010 this waste amounted to 31.9 million tonnes, and more than 8 million tonnes of it entered the ocean

The associate editor coordinating the review of this manuscript and approving it for publication was N. Ramesh Babu .

through multiple outlets. This is further highlighted in the interesting study of Ritchie and Roser [2], which reports that the amount of plastic waste - from tiny microplastics to larger macroplastics - floating on the surface of the world's oceans (excluding plastics found at depth or on the seafloor), reached a total of 268,950 tonnes in 2013.

These substances may directly or indirectly affect the life quality of aquatic organisms, as these latter can swallow them, risking suffocation, throat trauma, gastrointestinal obstruction, false satiation, impaired nutrition capacity, starvation, and even the death of some species [4]. According to the United Nations Environment Program, relatively high levels of very high concern substances (additives and absorbed toxins) are found by researchers in some plastic base production, such as textiles, clothing, and office/home magazines [5]. In the specific, microplastics derive from the fragmentation of these substances into smaller size fragments (less than 5 mm) and are considered as resistant contaminants in aquatic systems [6]. In fact, they have emerged as a pervasive environmental issue in recent years, as they have

the ability to infiltrate terrestrial and aquatic ecosystems, thus posing a significant threat to human health and the environment in its entirety [7].

Generally, microplastics can be categorized in primary and secondary microplastics [7]. The first ones are produced by the unintentional release of intermediate plastic feedstock (e.g., pellets, nurdles, etc) or by processes such as particulate emissions from industrial production, maintenance of plastic or plastic-based materials or release of dust and fibers [8]. Secondary microplastics can instead be produced by ultraviolet sunlight radiation [9] or by weathering [10], which cause the breakdown of large plastic into tiny fragments.

This work aims at improving the identification of primary microplastic derived from the wastewater of fibers washing processes. In fact, traditional methods of microplastic detection usually involve labor-intensive processes, often relying on chemical staining, visual identification and filtration, which are time-consuming and eventually prone-to-error techniques [11]. The advent of digital holography microscopy has instead provided a novel approach to the visualization and analysis of microplastics in water [12], which is further supported by the increasing exploitation of Deep Learning in several tasks, including microplastics detection and classification.

Within this framework, the release of the Holography Micro-Plastic Dataset (HMPD) [13] has given the opportunity to develop and test deep classification systems on a benchmark dataset. The results obtained from the presented preliminary experiments showed that popular deep learning models are able to obtain up to 83% of accuracy when using phase images as input; similar values can be obtained with amplitude data.

To enhance the accuracy and reliability of these findings, we present a novel color fusion schema that merges the amplitude and phase signals into a single image, which is then fed to the deep model. A detailed analysis of other possible fusion schemas demonstrates the importance of a proper data processing to maximize the performance. Additionally, leveraging the latest deep learning architectures significantly enhances accuracy, emphasizing the value of using a robust, pre-trained network for this task.

The paper is organized as follows: Section *Related works* gives an overview on the traditional and more recent approach to the task. Section *Method* shows the details of our approach, with the chosen hyper-parameters and the employed dataset described in *Experimental Setup* Section. The obtained performances are reported in *Results*, together with their discussion and the analysis of the other possible fusion techniques. Finally, some final considerations are provided in the *Conclusions* Section, highlighting some interesting possibilities for future works and improvements.

## II. RELATED WORKS

The detection and identification of microplastics gained significant attention in recent years, driven by the pressing need to understand and mitigate plastic pollution. Numerous

techniques have been developed and refined to achieve accurate and efficient microplastic analysis. Traditional optical techniques involve naked eye or stereo-microscope identification: even if scanning electron microscopy (SEM) can magnify a sample up to 500,000 times, allowing for the visualization of a particle and its surface structure, no detailed identification information on the polymer type can be obtained by all of these visual methods [14]. For this reason, these techniques are often combined with more elaborate methods, as in the case of energy-dispersive spectroscopy (which uses Xrays to activate the sample and identify its elemental composition); however, this is very costly and time-consuming [15]. The staining with Nile red is an improved semiautomated method which allows for the categorization of the particles according to their surface polarity, even if not revealing their chemical composition [16]. Visual inspections are often combined with the analysis of the burning properties and solvent resistance: these techniques recognize the polymer type and the associated additives at the same time, but they make the sample unavailable for other analysis due to the destructive nature of the method itself [17]. Another similar and fast analysis is based on a combination of thermogravimetric analysis and solid-phase extraction followed by thermal desorption gas chromatography-mass spectrometry [18]. Nevertheless, other than not giving information on the size and number of particles, the method is again destructive. Currently, the most effective and non-destructive methods in microplastic identification are based on the concept of exciting the sample and measuring the resulting spectra to identify the specific polymer material; that is, by IR spectroscopy and Raman spectroscopy [19]. However, the required instrumentation is expensive and the method is very time-consuming since the particles of interest for material identification need first to be visually isolated from the sample [20]. The work of Zarfl [21] reports a detailed analysis of the open issues and challenges derived from microplastic identification. Nevertheless, the impressive revolution of Deep Learning in fields like Computer Vision and Natural Language Processing has produced a sensible impact also on the microplastic classification task. Lee et al. [22] exploited a CNN model, boosted with a spatial attention mechanism, to distinguish between microplastic and natural organic matter by processing Raman spectra. Similarly, Zhang et al. [23] developed a one-dimensional CNN to perform microplastic classification on Raman spectra, demonstrating its effectiveness with respect to traditional machine learning approaches. Moreover, by exploiting a Mask R-CNN model [24], Han et al. [25] demonstrated the possibility of locating, classifying, and segmenting large marine microplastics; in particular, the proposed method is able to classify the microplastic sample as fiber, fragment, pellet, or rod.

As previously mentioned, digital holography microscopy has provided a great contribution to the visualization and analysis of microplastics in water samples [12]. Digital holography microscopy allows in fact for the acquisition of

both amplitude and phase information from the scattered light field, providing valuable insight into the three-dimensional structure of microplastics [26], making them distinguishable from other debris in water samples. For this reason, the digital holography-based microscopy approach is currently one of the most used techniques in the microplastic detection field. In fact, Bianco et al. [27] describe the most promising advances in digital holography and microplastic topics, while another of their works [28] describes a probing approach for microplastic identification through digital holography. In recent years, several works have exploited Deep Learning technologies in order to classify microplastics in digital holography images. In particular, Zhu et al. [29] employ an input compression block to reduce the number of input channels: together with a simple transfer learning technique, the proposed method is able to perform microplastic classification with pre-trained deep neural networks. Zhu et al. [30] describe a holographic classifier able to distinguish between microplastic and dust particles taking as input raw holograms, effectively providing a portable low-cost microplastics counting and classification tool. Moreover, Valentino et al. [31] propose fractal geometry features to enrich digital holography images, thus improving the classification accuracy of microplastics with respect to microplankton. Finally, Akkajit et al. [32] conduct a comprehensive analysis of different convolutional neural networks for digital holography imaging classification, focusing on the wastewater treatment plants setting.

### III. METHOD

This study aims at improving the classification accuracy of deep neural networks for microplastic recognition. Our approach focuses on analyzing amplitude and phase images obtained through digital holography techniques, enabling the extraction of microplastic samples from wastewater generated by washing synthetic clothes.

In general, these methodologies provide a complex image containing both the amplitude and the phase information of the sample: the complex field of the hologram is reconstructed through numerical techniques such as the Fourier or Fresnel transforms, and then the amplitude and phase images can be obtained by separating the respective components from the complex signal. Please refer to the study presented by Schnars et al. [33] for an in-depth understanding of the holography image processing technique.

In the case of the dataset used in this work, the obtained signals can be visualized as two grayscale images, the amplitude and phase images, both containing essential information for the correct sample classification, as demonstrated by Cacace et al [13]. For this reason, they serve as valuable input for training and testing neural networks to further develop accurate and robust classification models. More details regarding the data acquisition and the dataset creation are provided in the *Experimental setup* section. It is worth mentioning that, since the amplitude and phase

images capture different properties of the scattered light, the information content they provide is not the same.

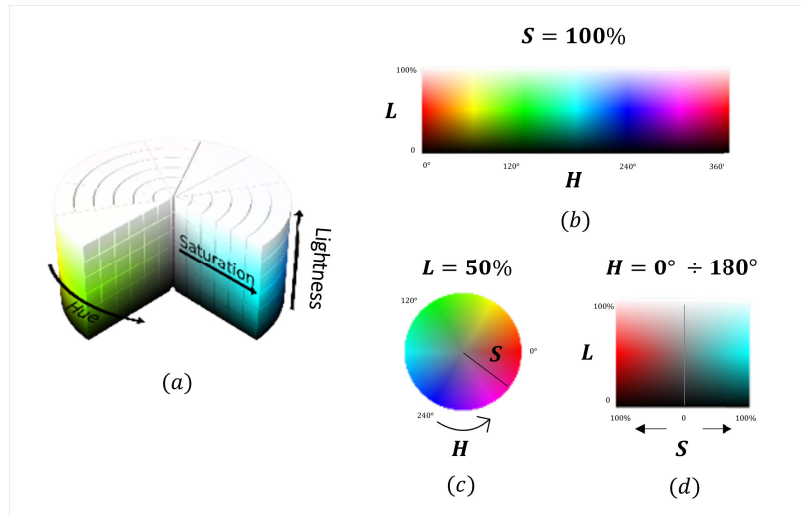
Starting from this assumption, we propose a color mapping schema capable of integrating the information content of both the amplitude and phase images into a single sample. This integration aims at maximizing the accuracy of a general deep neural network trained for microplastic classification. However, using the two one-dimensional images as input is not sufficient to fully define the typical RGB color space, on which deep networks are commonly pre-trained. To address this limitation, a possible solution involves assigning the amplitude and phase data to two of the three channels of an RGB image (e.g., the R and G channels) while setting the third one (e.g., the B channel) to a constant value or zero. Unfortunately, this transformation demonstrates to be sub-optimal for pre-trained networks on full-range RGB images, leading to the potential loss of valuable color information: that affects the overall performances of the networks, whose ability of feature extraction and classification relies instead on color cues, as demonstrated in table 1 of the *Results* Section.

To overcome this challenge, we propose the use of a cylindrical color space [34] to preserve the color information while integrating grayscale data. In particular, we adopted the Hue, Saturation, Luminance (*HSL*) color space [35], which encodes the full-spectrum color space into the three *Hue*, *Saturation* and *Luminance* channels. Specifically, the Hue values range from 0 to 360°, covering all the colours; the Saturation and the Luminance (responsible for the colour's saturation and brightness respectively) can be defined as real numbers between 0 and 1 (or in terms of percentage). The visual representation of the HSL color space is illustrated in Figure 1, offering an intuitive understanding of its components.

Given that we only have two informative data matrices, we set one of the three channels to a fixed value and re-modulate the others accordingly. Specifically, we rescale the grayscale amplitude image into the [0-359] range and assign it to the Hue channel. Conversely, the grayscale phase image is rescaled into the [0-1] range and assigned to the Saturation channel. Finally, the Luminance channel, responsible for balancing white/black levels in the image, is fixed at 0.5 (50%), which corresponds to the fully saturated pixel values. Formally, if we define the amplitude and phase images as one-dimensional matrices with dimensions  $m \times n$  pixels:

$$A_{m \times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad a_{ij} \in [0, 255]$$

$$P_{m \times n} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix} \quad p_{ij} \in [0, 255]$$



**FIGURE 1.** The Hue, Saturation, Luminance (HSL) model in its cylindrical form (a), with variations depicted in its three channels: (b) H and L with fixed maximum S, (c) S and H with L set at 50%, and (d) L and S variations in the Hue range of 0-180°.

The HSL image can be constructed as follows:

$$HSL_{m \times n} = [H, S, L]$$

$$H_{m \times n} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{bmatrix} \quad h_{ij} = \frac{360 * a_{ij}}{256}$$

$$S_{m \times n} = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \dots & s_{mn} \end{bmatrix} \quad s_{ij} = \frac{p_{ij}}{256}$$

$$L_{m \times n} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & l_{22} & \dots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \dots & l_{mn} \end{bmatrix} \quad l_{ij} = 0.5$$

The resulting HSL image is then converted into an RGB image with a standard HSL → RGB conversion procedure [36] to be fed to the deep neural network. As in the original HMPD work, we opted for a backbone network pre-trained on ImageNet: it undergoes a standard fine-tuning procedure on the HMPD dataset in which only the last fully connected layer is initialized from scratch, as it is the one responsible for the binary classification between microplastics and other debris categories. The proposed pipeline is represented in Figure 2.

Other possible choices of cylindrical color spaces can be used to merge amplitude and phase images; among all the possible alternatives, we tested the performances of the YCBCR color space as one of the most popular employed in the literature [37], where Y is the luminance component and CB and CR are the blue-difference and red-difference

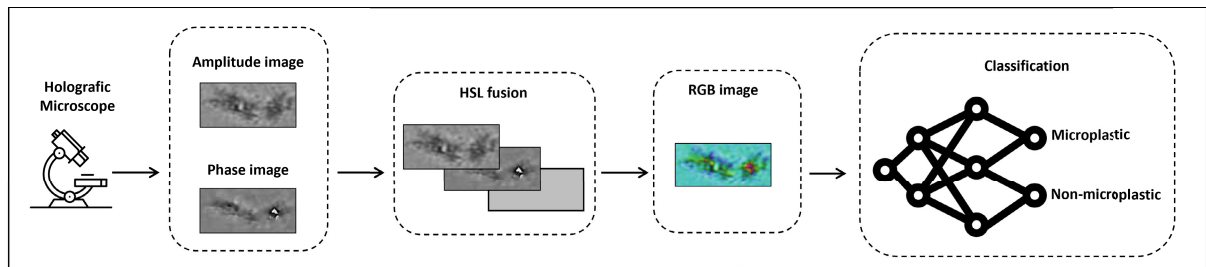
chroma components. In this case, we implemented the merge schema by assigning the amplitude and phase signals to the CB and CR channels respectively, and by fixing the Y channel (luminance) to a constant value equal to 127.

#### IV. EXPERIMENTAL SETUP

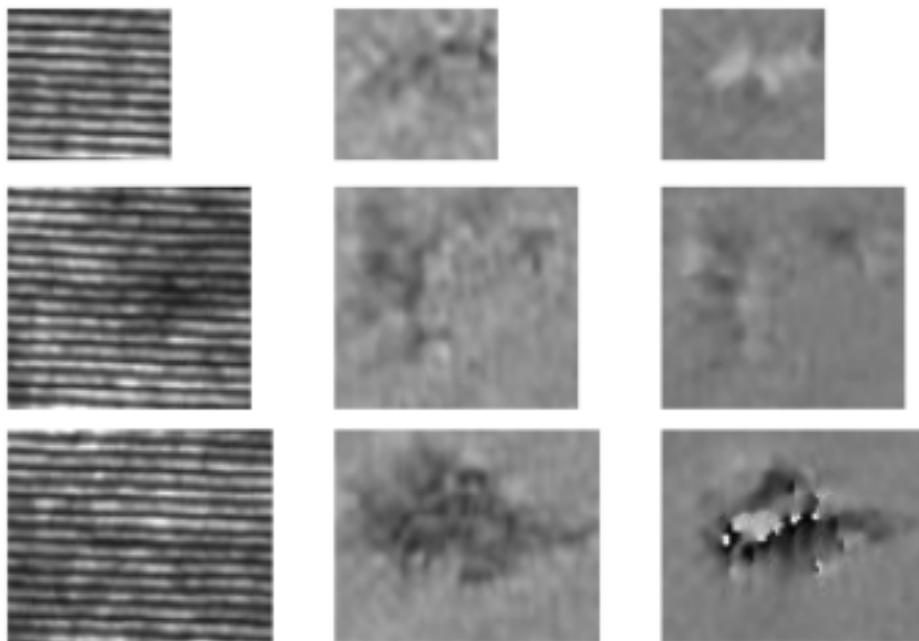
As previously mentioned, the benchmark dataset used for our experiments is the Holography Micro-Plastic Dataset (HMPD), which is publicly available at the [GitHub repository](#) along with the benchmark code. On this basis, we developed our benchmark code, available at this [GitHub repository](#).

The dataset comprises microplastic samples extracted from wastewater collected during a standard washing process of 100% polyester t-shirts. These wastewater samples underwent a multi-step filtration procedure before being analyzed using a low-cost Holographic Microscope. This technique facilitated the recording of the interaction between the scattered light from the sample and a reference clean wave, resulting in the generation of amplitude and phase images. Despite being provided in the dataset, the raw images were not utilized due to their low accuracy, as identified by the authors.

Specifically, two different wastewater samples were passed through the microscope's microchannel, leading to the acquisition of two raw video sequences. These sequences were segmented, and each frame (hologram) underwent a standard demodulation in the Fourier domain to reconstruct the amplitude and corresponding phase images. An object detection algorithm was then applied to the phase component to detect flowing objects, which were extracted as stand-alone patches and stored in a database with metadata for labeling. This latter procedure was performed by two experts in the environmental and optical fields, categorizing the samples as microplastic and non-microplastic. Samples with



**FIGURE 2.** Experimental workflow: the amplitude and phase images from the HMPD Dataset undergo HSL fusion to convert them into RGB images using the three channels. These RGB images are then classified as microplastic or non-microplastic using deep neural networks.



**FIGURE 3.** Microplastic samples extracted from the HMPD dataset, presented at their original resolution. The first column displays the sample in raw image format, while the second and third columns show the corresponding amplitude and phase images, respectively.

mismatched label categories were considered as possible-microplastics. For further details on the dataset creation, please refer to the original paper [13].

The HMPD dataset consists of 3293 positive (microplastic category) and 3293 negative samples, randomly selected from the non-microplastic category to ensure a balanced dataset. In Figure 3 three microplastic samples with the corresponding raw, amplitude and phase images are reported. The code provided by the authors includes a folding procedure that randomly splits the elements using a shuffling procedure, enabling the creation of five different training/testing splits, with an 80% - 20% proportion.

In their experiments, the authors tested some of the well-known deep neural architectures using the aforementioned patch images resized to  $80 \times 80$  pixels. The models were trained for 25 epochs with a learning rate of 0.00001 and a batch size of 32, with data augmentation employed to reduce overfitting. Specifically, the images were resized to  $100 \times 100$  pixels and then randomly cropped to the

desired  $80 \times 80$  final shape. Additionally, random rotations within a range of  $\pm 90$  degrees and random horizontal flips were applied. To ensure a fair comparison, we followed the same training/testing procedure and chose the same hyperparameters and resolutions. However, due to slight performance variability between different runs, we report the mean metric values over five different runs across the considered splits.

We tested the three deep neural network backbones employed in the original HMPD paper [13]: AlexNet [38], VGG11 [39], and ResNet18 [40], to demonstrate the positive effect of the HSL color schema on their performances. Moreover, we tested three additional models, namely DenseNet121 [41], ResNeXt50 [42], and Vision Transformer [43], with further increases in the final accuracy.

The AlexNet architecture revolutionized the field of image classification by demonstrating the effectiveness of deep convolutional neural networks on large-scale datasets. Despite being outdated compared to state-of-the-art models,

AlexNet remains widely used in the scientific community due to its groundbreaking performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), showcasing the potential of CNNs for image classification tasks. Comprising eight layers, including five convolutional layers and three fully connected layers, AlexNet utilized Rectified Linear Units (ReLU) activations and dropout regularization to mitigate overfitting, totaling 62.3 million trainable parameters.

The VGG11, a member of the VGG series, is a very popular network valued for its powerful pre-trained-on-ImageNet visual features. With five convolutional layers and three fully connected layers, VGG11 boasts 133 million trainable parameters. It maintains a fixed convolutional stride of one pixel, preserving spatial resolution and feature richness, and utilizes a softmax-activated output layer for efficient representation learning. VGG11 facilitates efficient transfer learning, making it a preferred choice for image classification tasks.

The ResNet18, belonging to the ResNet family, addressed the challenge of vanishing gradients through residual connections. With eighteen layers, including residual blocks, ResNet18 enables the training of deeper architectures without performance degradation. Residual connections allow gradients to flow directly through the network, facilitating smoother optimization and achieving state-of-the-art performance on various computer vision benchmarks. In the tested version, the eighteen deep layers are divided into four residual blocks.

Both DenseNet and ResNeXt are enhanced versions of the ResNet model. DenseNet emphasizes dense connectivity by adding an extra identity path which fosters more connections between lower and higher-level layers. This facilitates gradient flow and feature reuse, leading to efficient memory usage and enabling the construction of deeper networks without a proportional increase in parameters. DenseNet121, with 121 layers divided into a single initial convolutional layer, four dense blocks, and a final fully connected layer for classification, excels in image classification and transfer learning scenarios, leveraging dense connectivity for effective feature propagation.

ResNeXt50 extends the ResNet architecture by introducing a cardinality parameter, controlling the number of parallel paths within each residual block. By employing grouped convolutions, ResNeXt50 increases model capacity without significantly affecting the parameters total number, making it versatile and effective for image classification, object detection, and segmentation tasks.

Finally, the Vision Transformer (ViT) architecture represents a departure from traditional CNN-based approaches, leveraging self-attention mechanisms from transformer models. ViT treats images as sequences of non-overlapping patches and applies self-attention to capture global context effectively. Pre-trained on large-scale image datasets, ViT demonstrates effectiveness in both vision and language tasks, challenging CNN dominance in computer vision applications

and offering new avenues for cross-modal adaptation and innovation.

For all the considered backbones, the last fully connected layer (pre-trained on ImageNet) has been replaced by a new fully connected layer suitable for binary classification tasks and trained from scratch. The extensive tests have been made on a workstation equipped with an AMD Ryzen 7 5800X 8-Core Processor and a Nvidia RTX3060 graphic card.

The metrics used to evaluate the efficacy of our approach are the well-known Accuracy, Precision, and Recall metrics, defined as in (Eq. 1, 2 and 3):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$FPN = \frac{FP}{FP + TN} \quad (4)$$

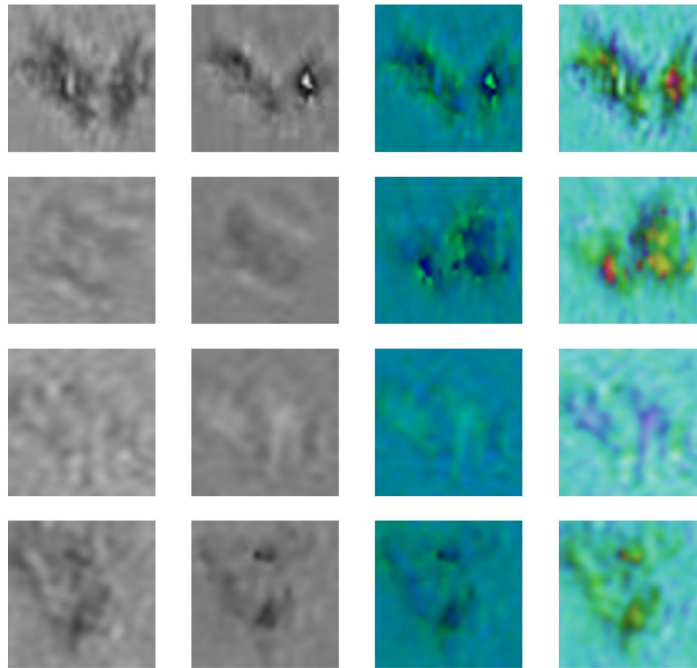
where TP, TN, FP, and FN represent the *True Positive*, *True Negative*, *False Positive*, *False Negative* respectively.

We also report the FPN value (Eq. 4), which allows to define the Receiver Operating Characteristic (ROC) curve as the plot of the Recall metric (also known as the True Positive Rate, TPR) against the False Positive Rate (FPN) at varying threshold values. The ROC curve illustrates the diagnostic ability of a binary classifier system, showing the trade-off between sensitivity (true positive rate) and specificity (true negative rate). In the *Results* section we report the ROC curve and the corresponding Area Under the ROC Curve (AUC), which quantifies the overall performance of the classifier across all possible thresholds. It represents the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. AUC values range from 0 to 1, where 1 indicates a perfect classifier, 0.5 indicates a classifier with no discrimination ability (equivalent to random guessing), and values below 0.5 indicate worse than random performance [44].

Figure 4 illustrates two samples categorized as microplastic and non-microplastic, showcasing their amplitude, phase, RGB, and HSL representations. Despite the availability of raw images in the dataset, the performance achieved on such data is notably low (70% accuracy). Hence, our analysis focuses solely on the results obtained using the amplitude and phase images.

## V. RESULTS

In this section, we compare the results of our proposed pipeline with those obtained using amplitude (A) or phase (P) images alone, as well as with other possible color mapping choices such as RGB and YCBCR. Table 1 presents the metrics values of accuracy, precision, and recall for all the combinations of backbones and input data. Results obtained in the HMPD original work are marked with an asterisk, and the best ones are highlighted in bold.



**FIGURE 4.** Samples extracted from the HMPD dataset after the 100 × 100 resizing, illustrating two microplastic samples (first and second rows) and two non-microplastic samples (third and fourth rows). Each column displays, sequentially, the amplitude, phase, RGB, and HSL images corresponding to the same sample.

**TABLE 1.** Performance assessment of four backbone architectures using the (HSL) color schema compared to amplitude (A) and phase (P) data, as well as other color schemas (RGB) and (YCBCR). Results from the original HMPD work are marked with an asterisk, while the best results are highlighted in bold.

Backbone and input	Accuracy	Precision	Recall	Backbone and input	Accuracy	Precision	Recall
AlexNet* (A)	82.4%	79.3%	87.5%	VGG11* (A)	84.0%	82.1%	86.8%
AlexNet* (P)	83.7%	83.4%	84.2%	VGG11* (P)	85.5%	82.9%	89.3%
AlexNet (RGB)	84.0%	83.5%	84.6%	VGG11 (RGB)	85.4%	83.9%	88.3%
AlexNet (YCBCR)	89.6%	89.6%	89.7%	VGG11 (YCBCR)	88.0%	86.4%	90.1%
AlexNet (HSL)	90.1%	90.7%	89.4%	VGG11 (HSL)	87.9%	86.6%	89.8%
ResNet18* (A)	80.9%	78.2%	85.7%	DenseNet121 (A)	88.8%	88.8%	88.8%
ResNet18* (P)	83.5%	82.6%	85.0%	DenseNet121 (P)	91.8%	92.0%	91.4%
ResNet18 (RGB)	82.8%	81.8%	84.4%	DenseNet121 (RGB)	92.8%	93.2%	92.3%
ResNet18 (YCBCR)	84.1%	83.8%	84.5%	DenseNet121 (YCBCR)	93.3%	93.0%	93.4%
ResNet18 (HSL)	84.6%	84.3%	84.8%	DenseNet121 (HSL)	93.8%	94.1%	93.5%
ResNeXt50 (A)	88.9%	88.2%	89.7%	Vit16 (A)	92.5%	92.8%	92.1%
ResNeXt50 (P)	91.1%	90.5%	91.8%	Vit16 (P)	94.0%	<b>93.4%</b>	94.8%
ResNeXt50 (RGB)	90.8%	90.3%	91.5%	Vit16 (RGB)	93.6%	93.1%	94.1%
ResNeXt50 (YCBCR)	91.0%	90.7%	91.4%	Vit16 (YCBCR)	94.2%	93.3%	95.1%
ResNeXt50 (HSL)	91.5%	91.0%	92.1%	Vit16 (HSL)	<b>94.7%</b>	93.2%	<b>96.4%</b>

The analysis reveals that the Vision Transformer model achieves the highest accuracy when combined with the HSL color space for amplitude and phase images. This configuration significantly improves upon the previous state-of-the-art result, surpassing the VGG11 with a single-phase image input by 9.2 points in accuracy.

Furthermore, when considering the accuracy produced by fixing the backbone and changing the color schema, a general performance improvement is achieved, especially if compared to the use of the phase image alone (even if containing more informative content than the amplitude one). For instance, with the AlexNet model, the baseline accuracy of 83.7% obtained with phase images improves by 0.3%

with RGB fusion, 5.9% with YCBCR colormap, and 6.4% with HSL schema. Similar trends are observed with other backbones such as ResNet18, DenseNet121, ResNeXt50, and Vision Transformer, where HSL outperforms other configurations except for VGG11, where YCBCR produces similar performances. These results underscore the effectiveness of the HSL color mapping in the enhancement of the accuracy thanks to (i) the full range of colors employed and (ii) to the exploitation of all the available information (given by phase and amplitude). While both the YCBCR and RGB outperform single amplitude or phase channels, they still fall behind the HSL in most cases.

The backbone choice significantly influences the performance, as evidenced by setting a fixed color fusion method and varying the backbone to assess the corresponding results. In particular, the Vision Transformer architecture achieves the best accuracy with respect to the other models even when considering the simple RGB method, thus confirming the importance of selecting a specific backbone to optimize the results. Among the tested convolutional neural networks, the DenseNet121 produces the best accuracy, with less than one point difference from the transformer model.

Regarding the other evaluated metrics, the suggested combination (Vision Transformer on HSL data) demonstrates a remarkably high recall value of 96.4%, with the second-best result at 94.8%. This indicates the model's ability to correctly identify the majority of microplastic samples while accepting a few false positives. Additionally, the obtained precision value of 93.2% ensures the model's proficiency in capturing relevant microplastic instances while minimizing false positive predictions, closely matching the best result obtained by the Vision Transformer network on phase data.

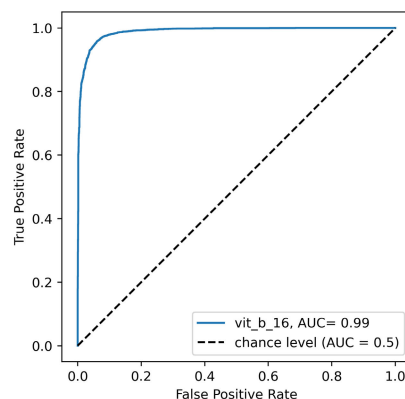
Figure 5 reports the ROC curve of the ViT model on HSL data. The extremely high AUC value of 0.99 confirms the validity of our proposed model in correctly discerning the two categories, as the value proximal to 1 indicates an optimal classifier. This further reinforces the robustness of our proposed model in accurately distinguishing between the two categories.

Moreover, to explore the impact of the constant channel choice in the HSL format, we conducted additional experiments in which we assigned a constant average value to the S channel instead of the L one, while setting the latter equal to the amplitude image. The resulting accuracy values generally demonstrate lower performance, ranging from a 5% accuracy difference on the AlexNet to a 0.4% difference with the Vision Transformer model. We conjecture that this effect arises from the enhanced image variability achieved by assigning the amplitude image to the S channel. This increased variability is particularly beneficial for lower-capacity models, leading to higher accuracies.

Regarding the training time, it is important to note the efficiency of the methodology. Given our cross-validation approach and the dataset subdivision into 5 folds, the training process averages approximately 3 hours for 25 epochs across each fold. That ensures robust model training while optimizing resource utilization.

It is also noteworthy mentioning the validity of the method towards the inference time: it demonstrates a considerable speed in producing the output on the workstation - with a single-image inference time of about 7 milliseconds (ViT) further reduced to 5 milliseconds with the ResNeXt and DenseNet models. Moreover, the conversion of the two input images into the HSL format adds a minimal overhead of 0.03 milliseconds.

We also evaluated the performance on a lower computational embedded device to assess the scalability and practical applicability of our method across a range of computing



**FIGURE 5.** ROC curve (blue line) of the Vision Transformer (ViT) model with HSL input data, indicating the corresponding Area Under the Curve (AUC) value. The black dotted line represents the performance of a classifier with chance-level accuracy.

environments, including those with limited resources. Specifically, we tested the models on an Nvidia Jetson Nano, a well-known embedded device for its interoperability in different settings (as we further demonstrated in other works [45]). In this case, we obtained an inference time of 55-58 msec for the Transformer model and 19-21 msec for the DenseNet121 model, thus confirming the method as well-suited for real-time deployment and applications where timely processing is necessary.

## VI. CONCLUSION

This study tackles the urgent issue of microplastic pollution in aquatic environments. These minute plastic particles have become a major environmental concern due to their widespread infiltration into terrestrial and aquatic ecosystems, posing threats to aquatic life, human health, and the environment as a whole. With the exponential increase in plastic production over the years, the durability of synthetic polymers, while advantageous in various applications, has also made them a significant source of pollution.

Our research aims at enhancing the detection and classification of microplastics by leveraging deep learning techniques. More in the specific, we propose a color fusion schema that simultaneously analyzes the amplitude and phase signals, thereby improving the accuracy of the classification. The experiments were conducted using the Holography Micro-Plastic Dataset, a publicly available benchmark dataset containing microplastic samples extracted from wastewater generated during the washing of synthetic clothes. Using a low-cost Holographic Microscope, the dataset authors recorded the interaction between scattered light and the samples, resulting in amplitude and phase images. Although these images are rich in information, they have distinct characteristics and offer different insights into microplastics, suggesting that their combined analysis could be more effective than analyzing them individually.

To address this, we propose a color fusion scheme based on the Hue-Saturation-Luminance (HSL) color space to



integrate the grayscale amplitude and phase signals into a single, informatively rich image. The HSL color mapping proves to be highly effective, significantly improving the accuracy if compared to traditional methods. We conducted several experiments with different deep neural network backbones, ranging from the original AlexNet (selected by the dataset authors) to the DenseNet121, ResNeXt, and the Vision Transformer, with the latter demonstrating the most outstanding performance, especially when combined with our proposed HSL color mapping schema. This result highlights the potential of color fusion methods to enhance microplastic classification performance.

To conclude, this work sets the stage for future advancements in microplastic research, aiming at supporting the comprehensive assessments of their environmental impact and developing strategies for monitoring and mitigation. Our research wants to contribute to the field of microplastic analysis, offering valuable insights for both environmental monitoring and human health protection. In future research, we plan to explore additional tasks such as microplastic detection and quantification, leveraging advanced computer vision techniques and deep learning architectures. Furthermore, we are collaborating with the experts to enhance the dataset quality and diversity, incorporating additional features and annotations; moreover, we will deepen the role of amplitude-phase image assignment with respect to the chosen color format, in order to better understand the different obtained performances and thus produce a more comprehensive overview of the phenomenon. Finally, we aim to further optimize our method for real-time deployment on low-powered embedded devices, considering factors such as computational efficiency, resource constraints, and scalability to support a broader range of applications in microplastic research and environmental monitoring.

#### AUTHOR CONTRIBUTIONS STATEMENT

Paolo Russo and Fabiana Di Ciaccio conceived the experiments, Paolo Russo conducted the experiments, Paolo Russo and Fabiana Di Ciaccio analyzed the results. Paolo Russo and Fabiana Di Ciaccio wrote and reviewed the manuscript.

#### DATA AVAILABILITY

The dataset analyzed during the current study is publicly available in the HMPD repository, [link](#)

#### ADDITIONAL INFORMATION

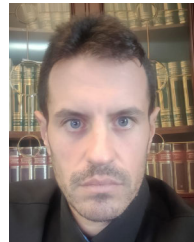
**Accession codes** Code publicly available at this [GitHub link](#).

**Competing interests.** The authors declare no conflict of interest.

#### REFERENCES

- [1] L. Van Cauwenberghe, L. Devriese, F. Galgani, J. Robbins, and C. R. Janssen, "Microplastics in sediments: A review of techniques, occurrence and effects," *Mar. Environ. Res.*, vol. 111, pp. 5–17, Oct. 2015.
- [2] H. Ritchie and M. Roser. (2018). *Plastic Pollution*. Our World in Data. [Online]. Available: <https://ourworldindata.org/plastic-pollution>
- [3] P. Li, X. Wang, M. Su, X. Zou, L. Duan, and H. Zhang, "Characteristics of plastic pollution in the environment: A review," *Bull. Environ. Contamination Toxicol.*, vol. 107, no. 4, pp. 577–584, Oct. 2021.
- [4] D. Eerkes-Medrano, R. C. Thompson, and D. C. Aldridge, "Microplastics in freshwater systems: A review of the emerging threats, identification of knowledge gaps and prioritisation of research needs," *Water Res.*, vol. 75, pp. 63–82, May 2015.
- [5] B. H. Desai, "14. United nations environment programme (UNEP)," *Yearbook Int. Environ. Law*, vol. 31, no. 1, pp. 319–325, 2020.
- [6] A. A. Koelmans, N. H. M. Nor, E. Hermesen, M. Kooi, S. M. Mintenig, and J. De France, "Microplastics in freshwaters and drinking water: Critical review and assessment of data quality," *Water Res.*, vol. 155, pp. 410–422, May 2019.
- [7] S. Sharma and S. Chatterjee, "Microplastic pollution, a threat to marine ecosystem and human health: A short review," *Environ. Sci. Pollut. Res.*, vol. 24, no. 27, pp. 21530–21547, Sep. 2017.
- [8] A. Anderson, A. Andradý, C. Arthur, J. Baker, H. Bouwman, S. Gall, V. Hildalgo-Ruz, A. Köhler, K. L. Law, and H. Leslie, "Sources, fate and effects of microplastics in the environment: A global assessment," GESAMP, Int. Maritime Org. (IMO), London, U.K., Tech. Rep., 90, 2015. [Online]. Available: <http://www.gesamp.org/publications/report-of-the-41st-session>
- [9] D. K. A. Barnes, F. Galgani, R. C. Thompson, and M. Barlaz, "Accumulation and fragmentation of plastic debris in global environments," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1526, pp. 1985–1998, Jul. 2009.
- [10] C. Arthur, J. Baker, and H. Bamford, "Proceedings of the international research workshop on the occurrence, effects, and fate of microplastic marine debris, September 9–11, 2008," *Nat. Ocean. Atmos. Admin. Tech. Memorandum*, Tech. Rep. NOS-OR&R-30, Jan. 2009.
- [11] W. J. Shim, S. H. Hong, and S. E. Eo, "Identification methods in microplastic analysis: A review," *Anal. Methods*, vol. 9, no. 9, pp. 1384–1391, 2017.
- [12] F. Merola, P. Memmolo, V. Bianco, M. Paturzo, M. G. Mazzocchi, and P. Ferraro, "Searching and identifying microplastics in marine environment by digital holography," *Eur. Phys. J. Plus*, vol. 133, no. 9, pp. 1–6, Sep. 2018.
- [13] T. Cacace, M. Del-Coco, P. Carcagní, M. Cocca, M. Paturzo, and C. Distante, "HMPD: A novel dataset for microplastics classification with digital holography," in *Image Analysis and Processing—ICIAP*, G. L. Foresti, A. Fusiello, and E. Hancock, Eds. Cham, Switzerland: Springer, 2023, pp. 123–133. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-031-43153-1\\_11#citeas](https://link.springer.com/chapter/10.1007/978-3-031-43153-1_11#citeas)
- [14] J. Li, H. Liu, and J. P. Chen, "Microplastics in freshwater systems: A review on occurrence, environmental effects, and methods for microplastics detection," *Water Res.*, vol. 137, pp. 362–374, Jun. 2018.
- [15] A. B. Silva, A. S. Bastos, C. I. Justino, J. P. da Costa, A. C. Duarte, and T. A. Rocha-Santos, "Microplastics in the environment: Challenges in analytical chemistry—A review," *Analytica Chim. Acta*, vol. 1017, pp. 1–19, Aug. 2018.
- [16] T. Maes, R. Jessop, N. Wellner, K. Haupt, and A. G. Mayes, "A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile red," *Sci. Rep.*, vol. 7, no. 1, p. 44501, Mar. 2017.
- [17] L. Mai, L.-J. Bao, L. Shi, C. S. Wong, and E. Y. Zeng, "A review of methods for measuring microplastics in aquatic environments," *Environ. Sci. Pollut. Res.*, vol. 25, no. 12, pp. 11319–11332, Apr. 2018.
- [18] E. Dümichen, P. Eisentraut, C. G. Bannick, A.-K. Barthel, R. Senz, and U. Braun, "Fast identification of microplastics in complex environmental samples by a thermal degradation method," *Chemosphere*, vol. 174, pp. 572–584, May 2017.
- [19] S. M. Mintenig, I. Int-Veen, M. G. J. Löder, S. Primpke, and G. Gerdt, "Identification of microplastic in effluents of waste water treatment plants using focal plane array-based micro-Fourier-transform infrared imaging," *Water Res.*, vol. 108, pp. 365–372, Jan. 2017.
- [20] L. Cabernard, L. Roscher, C. Lorenz, G. Gerdt, and S. Primpke, "Comparison of Raman and Fourier transform infrared spectroscopy for the quantification of microplastics in the aquatic environment," *Environ. Sci. Technol.*, vol. 52, no. 22, pp. 13279–13288, Nov. 2018.
- [21] C. Zarfl, "Promising techniques and open challenges for microplastic identification and quantification in environmental matrices," *Anal. Bioanal. Chem.*, vol. 411, no. 17, pp. 3743–3756, Jul. 2019.
- [22] S. Lee, H. Jeong, S. M. Hong, D. Yun, J. Lee, E. Kim, and K. H. Cho, "Automatic classification of microplastics and natural organic matter mixtures using a deep learning model," *Water Res.*, vol. 246, Nov. 2023, Art. no. 120710.

- [23] W. Zhang, W. Feng, Z. Cai, H. Wang, Q. Yan, and Q. Wang, "A deep one-dimensional convolutional neural network for microplastics classification using Raman spectroscopy," *Vibrational Spectrosc.*, vol. 124, Jan. 2023, Art. no. 103487.
- [24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [25] X.-L. Han, N.-J. Jiang, T. Hata, J. Choi, Y.-J. Du, and Y.-J. Wang, "Deep learning based approach for automated characterization of large marine microplastic particles," *Mar. Environ. Res.*, vol. 183, Jan. 2023, Art. no. 105829.
- [26] T. Cacace, S. Itri, M. Rani, S. Federici, L. Miccio, V. Bianco, M. Paturzo, and P. Ferraro, "Compact holographic microscope for imaging flowing microplastics," in *Proc. Int. Workshop Metrology Sea, Learn. Measure Sea Health Parameters (MetroSea)*, Oct. 2021, pp. 229–233.
- [27] V. Bianco, M. Valentino, J. Běhal, D. Pirone, S. Itri, R. Mossotti, G. Dalla Fontana, E. Stella, L. Miccio, P. Memmolo, and P. Ferraro, "Digital holography in microplastic identification," *Proc. SPIE*, vol. 12136, pp. 232–240, May 2022.
- [28] V. Bianco, P. Memmolo, P. Carcagni, F. Merola, M. Paturzo, C. Distanti, and P. Ferraro, "Probing micro-plastic items with coherent light enables their identification through digital holography," in *Digital Holography and Three-Dimensional Imaging*. Optica Publishing Group, 2019, pp. Th2A–2. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=DH-2019-Th2A.2>
- [29] Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holographic imaging and classification of microplastics using deep transfer learning," *Appl. Opt.*, vol. 60, no. 4, pp. 38–47, 2021.
- [30] Y. Zhu, C. H. Yeung, and E. Y. Lam, "Holographic classifier: Deep learning in digital holography for automatic micro-objects classification," in *Proc. IEEE 18th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2020, pp. 515–520.
- [31] M. Valentino, J. Běhal, D. Pirone, P. Memmolo, L. Miccio, V. Bianco, and P. Ferraro, "On the use of machine learning for microplastic identification from holographic phase-contrast signatures," *Proc. SPIE*, vol. 12621, pp. 216–223, Aug. 2023.
- [32] P. Akkajit, A. Sukkuea, and B. Thongnonghin, "Comparative analysis of five convolutional neural networks and transfer learning classification approach for microplastics in wastewater treatment plants," *Ecological Informat.*, vol. 78, Dec. 2023, Art. no. 102328.
- [33] U. Schnars, C. Falldorf, J. Watson, W. Jüptner, U. Schnars, C. Falldorf, J. Watson, and W. Jüptner, *Digital Holography*. Berlin, Germany: Springer, 2015.
- [34] A. Hanbury, "Constructing cylindrical coordinate colour spaces," *Pattern Recognit. Lett.*, vol. 29, no. 4, pp. 494–500, Mar. 2008.
- [35] N. A. Ibraheem, M. M. Hasan, R. Z. Khan, and P. K. Mishra, "Understanding color models: A review," *ARNP J. Sci. Technol.*, vol. 2, no. 3, pp. 265–275, 2012.
- [36] G. Saravanan, G. Yamuna, and S. Nandhini, "Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models," in *Proc. Int. Conf. Commun. Signal Process. (ICCCSP)*, Apr. 2016, pp. 462–466.
- [37] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative study of skin color detection and segmentation in HSV and YCbCr color space," *Proc. Comput. Sci.*, vol. 57, pp. 41–48, Jan. 2015.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
- [39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [42] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.
- [43] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [44] S. Vanderlooy and E. Hüllermeier, "A critical analysis of variants of the AUC," *Mach. Learn.*, vol. 72, no. 3, pp. 247–262, Sep. 2008.
- [45] P. Russo and F. Di Ciaccio, "Deep models optimization on embedded devices to improve the orientation estimation task at sea," in *Proc. IEEE Int. Workshop Metrology Sea, Learn. Measure Sea Health Parameters (MetroSea)*, Oct. 2022, pp. 44–49.



**PAOLO RUSSO** received the B.S. degree in telecommunication engineering from Università degli studi di Cassino, Italy, in 2008, and the M.S. degree in artificial intelligence and robotics and the Ph.D. degree in computer science from the Sapienza University of Rome, Italy, in 2016 and 2020, respectively.

From 2018 to 2019, he was a Researcher at Italian Institute of Technology (IIT), Turin, Italy. He is currently an Assistant Researcher with the Department of DIAG, AlcorLab, Sapienza University of Rome. He currently teaches interactive computer graphics. His main research interests include deep learning, computer vision, generative adversarial networks, and reinforcement learning. More information can be found on website <https://www.paolorusso.org> together with comprehensive publication lists.



**FABIANA DI CIACCIO** received the Ph.D. degree from the Parthenope University of Naples. She is the UNESCO Chair in environment, resources and sustainable development. She is currently an Assistant Professor (RTDa) with the Department of Civil and Environmental Engineering, University of Florence, and a member of the Geomatics for Environment and Conservation of Cultural heritage (GECO) Laboratory. Her research interests include attitude estimation methods based

on visual-inertial, computer vision and deep learning techniques, as well as cultural heritage preservation, climate change impact assessment, environmental monitoring, metrology, underwater photogrammetry, and 3D reconstruction techniques. She is part of the ISPRS WG II/7 for Underwater Data Acquisition and Processing. She served as the Guest Editor for *MDPI* journals and as a Special Session Chair for the International Workshop on Metrology for the Sea. She is one of the organizers of the First International Workshop on Computer Vision for Environment Monitoring and Preservation (ICIAP 2023) and the First Workshop on Machine Vision for Earth Observation and Environment Monitoring (BMVC 2023).

• • •

Open Access funding provided by 'Università degli Studi di Roma "La Sapienza" 2' within the CRUI CARE Agreement