



Vision-based techniques for automatic marine plankton classification

David Sosa-Trejo^{1,2} · Antonio Bandera¹ · Martín González¹ · Santiago Hernández-León²

Accepted: 27 February 2023 / Published online: 25 March 2023
© The Author(s) 2023

Abstract

Plankton are an important component of life on Earth. Since the 19th century, scientists have attempted to quantify species distributions using many techniques, such as direct counting, sizing, and classification with microscopes. Since then, extraordinary work has been performed regarding the development of plankton imaging systems, producing a massive backlog of images that await classification. Automatic image processing and classification approaches are opening new avenues for avoiding time-consuming manual procedures. While some algorithms have been adapted from many other applications for use with plankton, other exciting techniques have been developed exclusively for this issue. Achieving higher accuracy than that of human taxonomists is not yet possible, but an expeditious analysis is essential for discovering the world beyond plankton. Recent studies have shown the imminent development of real-time, in situ plankton image classification systems, which have only been slowed down by the complex implementations of algorithms on low-power processing hardware. This article compiles the techniques that have been proposed for classifying marine plankton, focusing on automatic methods that utilize image processing, from the beginnings of this field to the present day.

Keywords Marine plankton · Pattern recognition · Image processing · Plankton classification

✉ David Sosa-Trejo
david.sosa@uma.es

Antonio Bandera
ajbandera@uma.es

Martín González
martín@uma.es

Santiago Hernández-León
shernandezleon@ulpgc.es

¹ Department of Electronic Technology, University of Malaga, Malaga, Spain

² Unidad Océano y Clima, Instituto de Oceanografía y Cambio Global, Universidad de Las Palmas de Gran Canaria, ULPGC, Unidad Asociada ULPGC-CSIC, Canary Islands, Spain

1 Introduction

Plankton are tiny organisms living in the ocean. They are fundamental to the ocean's food chain, serve as an indirect indicator of water pollution (Suthers and Rissik 2019), and fix approximately 40% of the world's carbon (Falkowski 1994). When researchers started estimating the carbon cycle, they found it to be straightforward for terrestrial areas (Table 1, Riley (1944)), while the opposite was found for the ocean. The first approaches gave the ocean eight times the carbon fixation yield of land (155 Gt C yr^{-1} , Table 1.III, Rabinowitch (1945)). Later, this quantity was reduced to slightly less than the terrestrial value (15 Gt C yr^{-1} , Nielsen (1952)). Currently, this estimation lies between the cited previous values (approximately 69 Gt C yr^{-1} , del Giorgio and Duarte (2002)), but still presents considerable uncertainty. As an integral and necessary entity in the global carbon and nutrient cycles (Blaschko et al. 2005; Sieracki and Webb 1991) and a key element in regulating the planet's temperature (Blaschko et al. 2005; Hays et al. 2005), many approaches were proposed during the last few decades for accurately assessing the role of plankton in the global ocean.

The abundance, size, and taxonomy estimates of plankton organisms are of paramount importance for accounting for the role of living marine components in the flux, export, and sequestration of carbon to the interior of the ocean. Autotrophic organisms uptake CO_2 for photosynthesis in the euphotic zone, allowing them to fix anthropogenic carbon into organic matter in the ocean. This carbon circulates through the trophic web of the ocean, and a small fraction is transported downwards through passive and active fluxes as well as physical mixing (Buesseler et al. 2007). Passive flux is related to the sinking of organisms and particles, while active flux is transport carried out by zooplankton and micronekton migrants by consuming carbon in the shallower layers of the ocean and respiring, excreting, egesting, and dying in the meso- and bathypelagic zones (Hernández-León et al. 2019, 2020). The carbon transported by planktonic organisms towards the deep ocean will remain there for decades or centuries, avoiding the CO_2 accumulation process in the atmosphere and avoiding a faster increase in the planet's temperature. Another important fraction of this carbon will be respired in the upper layers of the ocean, returning to the atmosphere. The role of phyto- and zooplankton in this transport system is of main concern in climate change studies. To account for these fluxes, knowledge of the biomass and physiology of the ocean biota is required (e.g., Garijo and Hernández-León (2015)). Both parameters can be assessed through the abundance, sizes, and taxonomies of organisms. Since the direct counting, sizing, and classification of the microscopic plants and animals of the sea is quite laborious and time-consuming, oceanographers have relied on technology to more quickly quantify these organisms.

The first method employed a microscope to estimate the inorganic and organic suspended matter retained in a molecular filter (Goldberg et al. 1952). Since microscopic work is tedious, Jerlov (1955) employed a Tyndall metre, which presented a relationship between seawater scattering and the total surface area of the particles contained on it. Although the Tyndall metre readings had an accuracy of $\pm 10\%$ above that of a microscope, the predominance of small and transparent particles led to measurement problems, thus highlighting microscopy as a superior method.

After succeeding in counting blood cells (Mattern et al. 1957), scientists began to use the impulse Coulter counter (Model A) to count marine organisms (Hastings et al. 1962; Maloney et al. 1962). Organisms placed in an electrically conductive medium pass through a small aperture, producing a voltage drop depending on the organism's size.

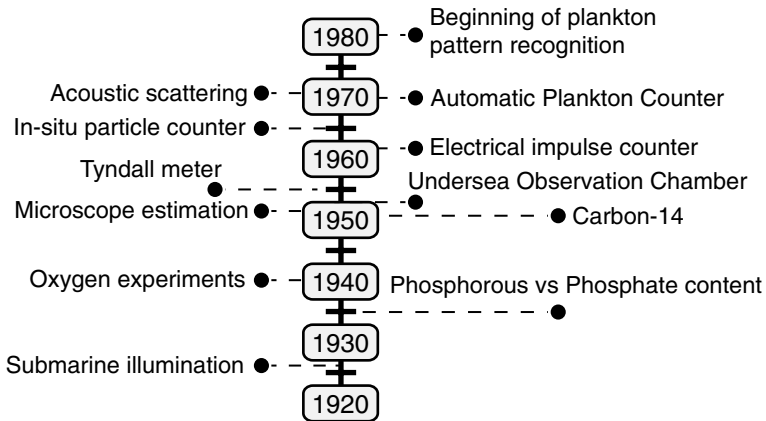


Fig. 1 Initial approaches for plankton quantification

Then, these impulses are amplified, recorded, and visualized. Organisms with diameters from 3 μm to 1 mm are counted and sized by using different aperture sizes (Sheldon and Parsons 1967b), with densities ranging from 50 to 100,000 cells/ml (Hastings et al. 1962). This was a significant advance in accounting for the abundance of phytoplanktonic cells, in terms of not only accuracy but also counting speed. The posterior Coulter counter version (Model B, Sheldon and Parsons (1967a)), jointly with an automatic cell-size distribution plotter (Model J), even allowed researchers to measure aggregate forms (i.e., chain-forming diatoms, Parsons (1965)). Despite its accuracy and speed, this method counts all falling particles in the same size range (Rehnberg et al. 1982) without regard to live or dead cells, fragmented cells, or debris. Maddux and Kanwisher (1965) developed an in situ particle counter by applying the same Coulter counter principle but installing it at the cod end of a tow-net. Later, other systems were developed, such as the automated plankton counter (Cooke et al. 1970), which are capable of sizing larger particles than those handled by the Coulter counter. Its operating mode is as follows: an organism's image is projected on an array of photosensors while a fluid stream carries it; when the path of light to the first photosensor is blocked, an impulse sequence initiates, finishing when the first sensor is again unblocked, indicating the organism's passage. This instrument was practical for counting and sizing organisms, eclipsed only by sample clogging (Fulton 1972). In 1971, acoustics was incorporated into plankton quantification (Beamish 1971). However, until now, captured samples were needed to know what the echo caught; this idea is known as ground truthing (McClatchie et al. 2000). This method is helpful for spatial distribution determination but not for identification.

Significantly, none of these techniques found general acceptance by the scientific community, probably because they were not explicitly developed for dealing with this specific scenario (Jeffries et al. 1981) and could not discriminate between taxonomic groups (Jeffries et al. 1984). Furthermore, all these methods were developed to process water samples. In contrast, in 1953, Nishizawa et al. (1954) captured in situ underwater photographs of suspended matter and plankton using a so-called undersea observation chamber (Fig. 1). Therefore, the authors observed a more common distribution of larger particles than those reported in the literature (Goldberg et al. 1952; Jerlov 1953). They considered that the water sampling process could damage or disintegrate these

large particles. Other authors agreed with this (Riley 1963; Mullin 1965), noting how natural aggregations could be broken up during sampling or rendered unidentifiable and biased when using standard preservation techniques (Murphy and Haugen 1985; Ortner et al. 1981; Zarauz and Irigoien 2008), “*which probably accounts for the fact that this phenomenon passed largely unnoticed until the advent of direct undersea observation*” (Riley 1963). Hence, in situ water photography presented an excellent sampling approach compared to traditional net-sampling techniques, demonstrating the advantage of sampling the fragile taxa that would be damaged in other circumstances (Olney and Houde 1993; Tiselius 1998).

Since the 1980s, many plankton imaging systems have emerged, from laboratory to in situ underwater video systems (Table 1, Lombard et al. (2019)). These systems have produced a massive backlog of unlabelled plankton data, which, in conjunction with the limited availability of human taxonomic expertise, represents a thread for future scientific studies (Simpson et al. 1992; Culverhouse et al. 2006; MacLeod and Association 2007). This review aims to describe and compare all the image classification algorithms that have been implemented to date for marine plankton. In this way, we consider that this work may be helpful for those who want to develop new image classification algorithms for application to marine plankton in the future.

The rest of the paper is organized as follows. Section 2 briefly defines plankton. Section 3 describes the selected criteria for this review. Section 4 divides the work developed in the selected papers grouped into six categories depending on their classification techniques. Table 1 presents a summary of these reviewed papers. In Sect. 5, the conclusion and future research directions are discussed.

2 A look at plankton

Plankton drift into the ocean in swarms, a concept known as patchiness in biological oceanography (Folt and Burns 1999). These small organisms range in size from tenths of microns up to centimetres and are categorized into phytoplankton (mainly autotrophs, acting as plants) and zooplankton (mainly heterotrophs, acting as animals). Figure 2 shows snapshots of plankton that belong to four valuable datasets found while producing this review. A vast and diverse image dataset is vital for developing new classification procedures.

Figure 2A shows a phytoplankton sample (dinoflagellate, ceratium) from the dataset released by the Woods Hole Oceanographic Institution (WHOI). The WHOI-Plankton dataset was one of the first public databases collected in situ with the Imaging FlowCytobot (IFCB) at Martha’s Vineyard Coastal Observatory, Massachusetts (Sosik 2015). It contains more than 3.5 million labelled images falling into 103 categories.

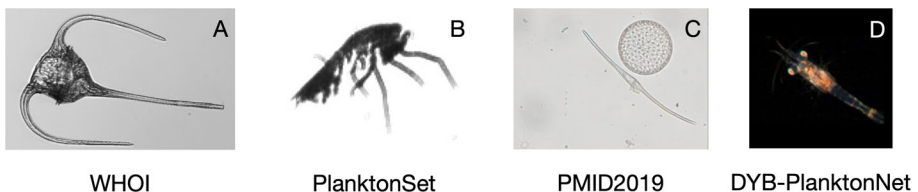


Fig. 2 Samples from four plankton datasets

Figure 2B shows a zooplankton sample from the PlanktonSet dataset (crustacean, amphipod) released by Oregon State University's Hatfield Marine Science Center. This dataset was initially offered to the National Data Science Bowl (NDSB) competition, collected with the In Situ Ichthyoplankton Imaging System (ISIS) in the Straits of Florida (Cowen et al. 2015). It contains 30,000 labelled images falling into 121 categories, plus a test set of unlabelled images.

Figure 2C shows an image containing two phytoplankton from the PMID2019 dataset (diatom, *coscinodiscus* - rounded one, and *dinoflagellata*, *ceratium fusus* - elongated one) released by the Ocean University of China, Qingdao. This was the first available high-resolution phytoplankton colour dataset captured with an Olympus BX53 fluorescence microscope from preserved samples collected in Qingdao Jiaozhou Bay, Shandong (Li et al. 2019). It contains over 10,000 labelled phytoplankton images from 24 categories.

Figure 2D shows a zooplankton sample from the DYB-PlanktonNet dataset (crustacean, shrimp-like) released by the Chinese Academy of Sciences, Shenzhen Institute of Advanced Technology. This dataset was collected with an innovative in situ colour imager installed on a buoy in the South China Sea near Shenzhen city (Li et al. 2021b). It contains over 46,000 labelled plankton images from 90 categories.

3 Article selection criteria

The reviewed papers were selected based on the following exclusion criteria.

Recent papers classifying low-quality captured images were discarded because they lacked details and were blurred or out of focus, making their classification processes even more complicated and distracting from the aim of this paper.

Classification studies using flow cytometry (FC) data were discarded because they employed fluorescent and light scattering characteristics for classification instead of a specimen's picture (McKinnon 2018). However, we included papers that classified images acquired with the Imaging FlowCytobot or the FlowCam; these instruments are based on the FC principle but add video technology to produce high-resolution photos.

Papers classifying images with specific features related to their acquisition methods, such as fluorescence microscopy (Blackburn et al. 1998; Rodenacker et al. 2006; Ng et al. 2017), polarized microscopy (Tiwari and Gallager 2003), or luminous events of bioluminescent plankton (Kocak et al. 1999), were discarded. These classification procedures employ features that are different from those extracted from a specimen's image.

In addition, articles with low correct classification rates were also excluded (<60% accuracy), as were articles with automatic > 99% classification rates (e.g., Loke et al. (2004)). Most recent articles with similar procedures to those used in the past were not considered (e.g., the work of Zhou et al. (2008) was similar to that of Liu et al. (1994) and Thonnat and Gandelin (1988)). Articles with equivalent procedures but achieving meagre accuracy improvements were discarded (e.g., Kramer et al. (2011) improved their accuracy by as much as 2.1%), and articles categorizing fewer than four classes were also discarded.

This study reviewed novel techniques in plankton classification, categorizing black and white or greyscale images, from their beginnings until today.

4 Image classification chronology

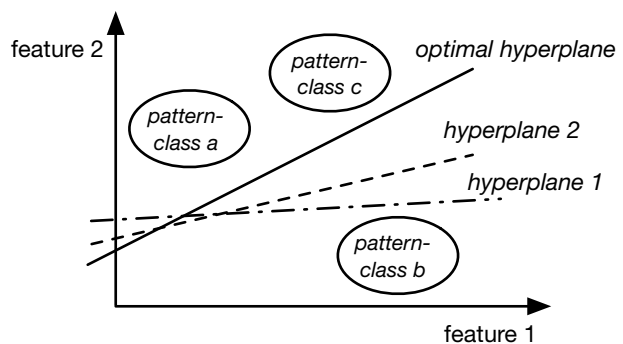
In 1970, Health Sciences established the techniques used in the forthcoming automated plankton identification systems, with a special-purpose computer extracting a set of parameters describing blood cell features for later identification using pairwise separation (Ingram and Preston 1970). Due to the inability to perform automatic plankton identification with these methods (see the introduction), Fawell (1976) introduced a similar technique using image analysis equipment (Quantimet 720, IMANCO (1970)) to extract morphometric features from plankton samples for later use in an image classifier. This strategy revealed the path for future plankton identification works until the advent of deep learning, where feature extraction and classification are intimately tied together.

4.1 Classification via linear discriminant analysis

Linear discriminant analysis (LDA) is a popular supervised classification technique that assumes a normal or Gaussian distribution for data points and identical covariance matrices for each class. It is also known as Fisher's LDA, as Ronald A. Fisher developed it in the 1930s. LDA aims to find the projection hyperplane that maximizes the distance between the projected means of the given classes and minimizes the variance within the classes; this can be stated as a generalized eigenvalue problem in a multiclass scenario. If c is the number of training classes, we can find $c - 1$ discriminatory directions that separate the c classes as much as possible (the largest eigenvectors, Fig. 3).

Uhlmann et al. (1978) designed a pattern recognition system for phytoplankton. For image acquisition purposes, they built a custom camera device (vidicon) to scan either phytoplankton photographs or species preserved in formalin. After performing contrast enhancement, the control unit transferred the scanned images to a computer (Robotron KRS 4200), which processed the patterns into a 2-D power frequency spectrum, defining a position in the polar space that was independent of translations and rotations. Two years later, Schlimpert et al. (1980) published the details of this work. To construct a linear classifier, they prepared a set of training patterns comprising five classes, which were iteratively presented to the classifier while tuning its parameters. As this technique produced unsatisfactory results, they combined two interlacing classes, trained the classifier again with four classes, and separated these combined classes in a posterior second stage. They obtained a 93.2% correct classification rate when categorizing the five classes. Even with this remarkable result, they expected that a better classification accuracy could be obtained

Fig. 3 Iterative construction of an optimal hyperplane in a 2-D feature space (Schlimpert et al. 1980)



if the scan resolution was improved. The 2-D power frequency spectrum processing step took approximately 10 s per pattern, and the classifier converged between 3 and 170 s, depending on the class. Nevertheless, the authors were aware that higher speeds could be attained by using parallel processing. Given the resolution of the scanned images shown in their paper, this work was a remarkable effort at the time. However, LDA is not the most appropriate method when some input variables are correlated, as shown in this work.

Jeffries et al. (1980) also employed LDA but instead categorized zooplankton. The idea was to classify species into ecologically and taxonomically meaningful groups. To this end, they identified which species were viable for classification according to morphometric relations. They employed two datasets: photocopied plates of species obtained from published literature and contours of preserved specimens traced by hand on acetate sheets; the first was captured with a vidicon, and the second was obtained with the Bausch and Lomb QMS system. The first dataset was classified into six shape categories using LDA by the jack-knife procedure, yielding a 93.1% correct classification rate. The second dataset was classified into 19 major species using the LDA-based nonpooled covariance procedure, reaching a 97.9% correct classification rate due to the combination of four morphometric relations. This work marked the beginning of automatic zooplankton image classification, achieving an 85% time reduction compared to manual processing in routine laboratory work.

Four years later, Jeffries et al. (1984) built a custom vidicon system to capture better plankton images. As a novelty, six satellite microprocessors were employed for feature extraction. These microprocessors worked in parallel under a central computer, obtaining an 89% correct classification rate when categorizing eight groups with the same five features as those in their previous work (Jeffries et al. 1980). Katsinis et al. (1984) improved this classification process by employing eight satellite microprocessors to extract nine morphometric features, obtaining a 92% correct classification rate in less than half the time required by Jeffries et al. (1984). These studies employed custom edge detection techniques to erase all zooplankton antennae and swimming legs, which led to problems when performing feature extraction. Although the plankton group classification results were promising, species identification was still impossible due to the poor quality of the input images. Additionally, it was necessary to manually check that no organisms overlapped before scanning. These works are noteworthy due to the star architectures of their central computers and satellites (the microprocessors), demonstrating how easy it is to implement an LDA classifier. However, as stated before, an LDA classifier may not be the most appropriate classifier when similar plankton species exist.

4.2 Classification via hierarchical clustering

Hierarchical clustering (HC) is an unsupervised classification algorithm that organizes a group of unlabelled data points into a tree-shaped structure (called a dendrogram). The dendrogram can be built in a bottom-up manner, in which all data points are initially considered as single clusters that are iteratively merged until only one cluster remains, or in a top-down manner, in which all data points initially belong to a single cluster that is recursively split until each cluster includes only a single data point. These approaches are named agglomerative and divisive HC, respectively (Fig. 4). HC gained notoriety after the publication of the "Numerical Taxonomy" book (Sneath and Sokal 1973), where the authors presented the advantages of HC in the field of biology and its application to organism classification.

Chehdi et al. (1986) from ENST Bretagne collaborated with the IFREMER in Brest using an agglomerative HC approach to classify 16 groups of zooplankton (Fig. 4A). A camera installed on a microscope captured images, and Freeman’s chain code was used to encode the contours. The appendages were removed to focus on bodies, separating those that seemed to overlap. They tested this system on 320 specimens by employing ten morphometric features; the algorithm could only find specimens falling into eight biological groups. Although this result was not the best, biologists liked how the approach gathered specimens, but improvements in image quality were mandatory for better classification. Six years later, Chehdi and Coquin (1992) optimized this method to classify 14 groups of zooplankton. As a novelty, they preprocessed the input images to resolve their problems concerning shadows, lights, and high-frequency noises. Seven features, five morphometric and two inertial features, were extracted to orient the specimens. They successfully classified 280 specimens into 14 groups, obtaining a 78.8% correct classification rate. As a drawback, this method required manual supervision to control the grouping at each aggregation level.

Thonnat and Gandelin (1988) applied a divisive HC method to classify Mediterranean zooplankton and their developmental stages for the first time (Fig. 4B). After performing photogram digitization and organism detection, they extracted morphological, densitometric, external and internal parameters. Then, marine biologists defined 59 prototypes containing descriptions for each class and subclass of zooplankton. Finally, they proceeded classify the plankton with 66 conditional rules, where each image took 122 s to process. They tested this classifier with only 40 zooplankton images but provided no information on accuracy, stating that they were working on constructing a more extensive dataset. This plankton application resulted from extending their previous work (classifying galaxies with HC) to other areas. While applying HC to plankton classification seems interesting, it may not be the most appropriate method because it does not use prior knowledge about the relationships between the different species. However, the employed tree-like structure provides a valuable visual representation, helping scientists understand how different species are related to one another and enabling them to identify patterns.

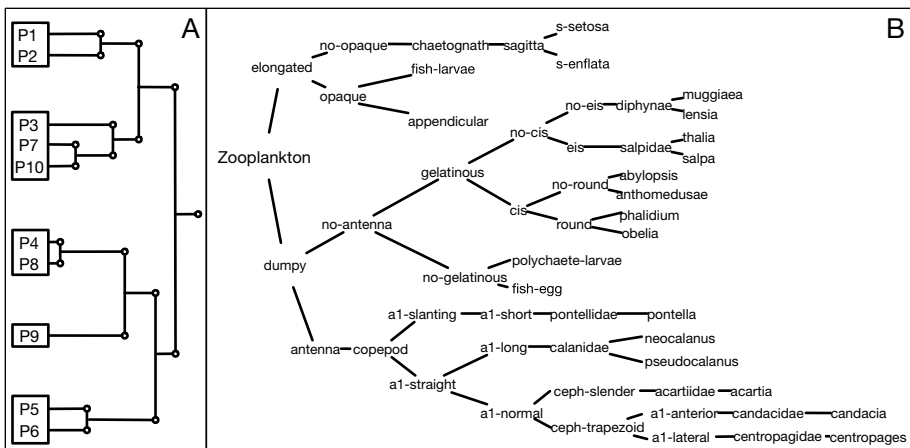


Fig. 4 HC methods: agglomerative (A) and divisive (B)

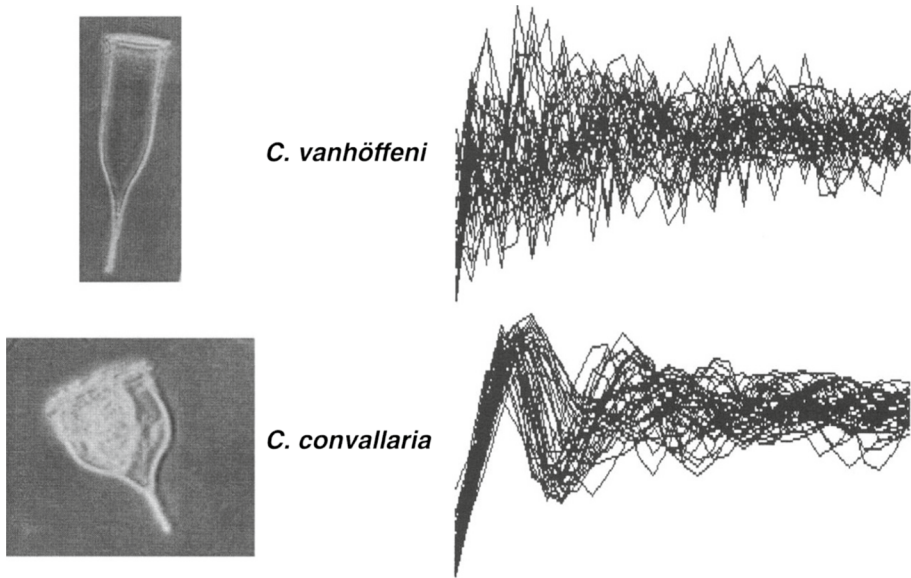


Fig. 5 Sixteen-bin FFT histograms for *C. vanhoeffeni* and *C. convallaria*. An ANN correctly classified 53% of the *C. vanhoeffeni* samples and 98% of the *C. convallaria* samples (Culverhouse et al. 1994)

4.3 Classification via artificial neural networks

To mimic the architecture of the human brain, artificial neural networks (ANNs) were introduced in the 1940s. ANNs are composed of artificial neurons. Each artificial neuron has inputs and generates a single output, which can be provided as an input to other neurons. These connections, such as the synapses in a biological brain, can transmit a signal from one neuron to another. A weight value increases or decreases the strength of the signal traversing a connection. A typical ANN organizes its neurons into several layers: one input layer and one output layer with hidden layers between them. The weight values are usually adjusted by minimizing the difference between the processed output of the ANN and the target output. This training procedure is the most relevant aspect of forming a reliable model.

Simpson et al. (1991, 1992) published the first work that applied a ANN for plankton identification. Three years later, Culverhouse et al. (1994) employed the same network but increased the number of categories from two to five congeneric species, which were difficult to classify even for taxonomists. The topology of the ANN was a 15-3-5 fully connected network (every neuron in one layer was connected to every neuron in the next layer). Backpropagation was used to tune the weight values of the ANN based on the error rate obtained in the previous iteration. The employed dataset was the same as that used by Williams et al. (1994), which was comprised of digitized photomicrographs that were validated by six experts into two subdatasets: one for training and another for testing. After generating Fourier spatial frequencies as features, they ran several trials with a randomized dataset size. The minimum obtained mean test error was 23% on unseen images. Species that were difficult to classify by experts were also difficult to classify with the ANN due to how bad the Fourier components of these species were, as clearly seen in

Fig. 5. This motivated the authors to look for other preprocessing techniques in conjunction with adjustments to the Fourier technique and their ANN. The authors knew that there was room for improvement because Williams et al. (1994) proved that this dataset was discriminable with 95% confidence by conducting a multivariate statistical analysis on the Fourier-transformed data. This work is exemplary for understanding how ANNs can be sensitive to data quality, performing poorly when the input data are noisy or contain outliers.

Ellis et al. (1994) also employed the same technique used by Simpson et al. (1991, 1992) to classify four plankton species. As a novelty, they combined the Fourier spatial frequencies with the internal textures of the plankton bodies extracted using the Gabor function. They achieved a 73% accuracy rate using these combined features, but the accuracy was only 69.8% or 43.7% when using only the Fourier-transformed spatial frequencies or the texture features alone, respectively. This ground-breaking work was the first to look at the advantage of combining both types of features in the field of plankton classification. However, as shown in their paper, image preprocessing is crucial for attaining high accuracy, making this technique very time-consuming for plankton classification and other applications.

Culverhouse et al. (1996) developed an automatic system for classifying 23 species of dinoflagellates identified as having negative impacts on the aquaculture industry. They first digitized the specimens archived at the Plymouth Marine Laboratory and fresh samples collected by the Spanish Institute of Oceanography. Afterwards, these samples were labelled by experts to create a database. Then, they determined the edges using the Sobel operator, where 60 features were extracted, including shape and texture features. Later, they compared the classification performance of two ANNs with that of two classic multivariate statistical techniques. The radial basis function (RBF) ANN performed the best with an 83% classification accuracy compared to that of taxonomists, who achieved an 85% rate. The authors highlighted these results because the classifiers were trained and tested from field-collected specimens instead of cultured specimens. While this system could not handle overlapping specimens, detritus-contacting samples did not appear to affect the accuracy of classifier because they were much smaller than plankton. This system evolved slightly, and in 2000, the software was designated as dinoflagellate categorization by an ANN (DiCANN) (Culverhouse et al. 2000). Although taxonomists slightly outperformed the ANN, this study evaluated how ANNs were superior to other discriminant methods, even to taxonomists, given that their better accuracy could fade away with psychological factors such as short-term memory, fatigue, boredom, recency effects, and positive bias. The potential sample analysis time reduction from human taxonomists to an ANN was from 120 to 5 min. Additionally, this work confirmed that data preprocessing is important for ANNs; therefore, the authors had to develop a specific preprocessing technique to manage debris.

Tang and his colleagues (Tang and Stewart 1996; Tang et al. 1998) produced a pattern recognition system to classify the massive quantity of images collected with the Video Plankton Recorder (VPR) (Davis et al. 1992). First, a video processing system rejected objects that were out of focus. Next, the system extracted features with three methods: moment invariants, Fourier boundary descriptors, and granulometric features; the three techniques grabbed shape information, while the latter method also acquired texture information. Later, the authors removed redundant information by reducing the feature vectors using the Karhunen-Loeve transform (KLT). Finally, they applied the supervised learning vector quantization (LVQ) ANN classifier to six classes of plankton. Due to the time-consuming nature of the classifier, they used a parallel training strategy. Regarding the results, they focused on which feature or feature combination gave the best accuracy instead of

studying which were the best network parameters, obtaining a 92.2% accuracy on the test set using the combined feature vector. This work demonstrated the benefit of using granulometric features alone or in combination with other features, as shown in Table 1 of their paper, which helped counteract the noise, occlusion, and variability of the dataset. It is also worth noting the innovative parallel strategy used to reduce the training time of the LVQ algorithm.

Ellis et al. (1997) developed an alternative method for classifying six phytoplankton species. It combined the results of different individual classifiers into one result with an arbitration algorithm; this approach was established based on the idea that a group of experts produces a more accurate classification than an individual expert. They named their technique "classification by committee". First, they described the image textures with a Gabor filter, which enabled the input image to be segmented into a background and foreground. Then, five different shape and texture feature vectors were extracted from the foreground and individually employed to train five ANNs, obtaining five different classification performances. These five different results were combined with three algorithms based on several criteria: the majority of votes, the sum of votes, and the most confident network, where each network had a single vote. Among the three criteria, the most confident network combination offered a maximum accuracy of 84%. Finally, the authors compared these "classification by committee" techniques with another method called "classification by collective machines", which combined the five different obtained results with an ANN instead of criteria. This latter method slightly outperformed "classification by committee" (70.3% versus 68.7% accuracy, respectively). In all cases, this work showed that combining networks' results with a criterion or an ANN resulted in higher accuracy than picking the best single ANN, but this was only true when the individual ANNs had comparable accuracies. This work tried to increase classification performance based on perspectives other than those used in previously reviewed papers. Although these techniques were innovative for plankton classification, the authors were inspired by prior works (Jacobs et al. 1991; Wolpert 1992; Battiti and Colla 1994).

Two years later, Dollfus and Beaufort (1999) applied position-normalized images directly on an ANN instead of using feature vectors. The authors employed free weights because neurons specialized better in specific places (in normalized images), increasing the probability of correct classification. Thus, they obtained an average correct classification rate of 86% when categorizing 13 species of calcareous nanoplankton. This classification system was named SYRACO2, proving helpful for natural forms with high intravariability. In addition, they found that algorithms performing position normalization needed to be more robust. Therefore, in 2004, they employed parallel ANNs (Beaufort and Dollfus 2004) in a system capable of categorizing nonposition-normalized images. Thus, the images were presented directly to the ANN (vision module), and in the case of weak identification, a second ANN (motor modules) performed up to five actions on each image (translation, rotation, dilatation, contrast, and symmetry) that helped to confirm or deny the initial weak identification result. This updated system achieved 91% accuracy in classifying 11 categories, with a speed of 7 s per image. It is notable that in this work, the authors attained a better accuracy by confronting the weak points of ANNs: the complicated pre-processing step and their effectiveness with correlated inputs. The direct image input process used by the network was similar to the procedure employed in a convolutional neural network (CNN).

Hu and his colleagues (Hu and Davis 2006; Hu 2006) presented a new system employing two classifiers: the first was an ANN using shape-based features (the same as in Tang and Stewart (1996); Tang et al. (1998)), and the second was a support vector machine

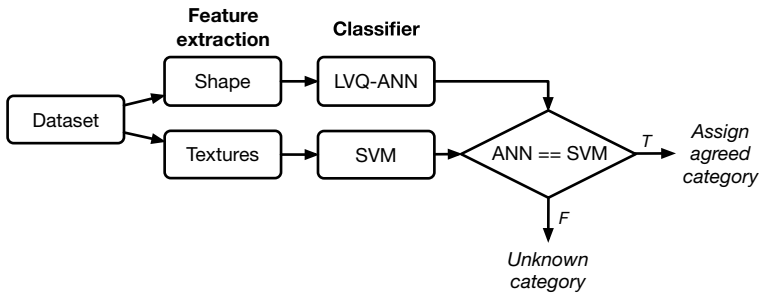


Fig. 6 Dual-classification system implemented by Hu and Davis (2006)

(SVM) using texture-based features (the same as that in Hu and Davis (2005)). Then, both classifiers' outputs were combined, resulting in a final correct identification if both agreed to the same class or an unknown result otherwise, as shown in Fig. 6. This system achieved 64.7% accuracy compared to the 61% accuracy attained by Davis et al. (2004) on the same dataset using an LVQ ANN classifier. The dataset was collected with the VPR and comprised seven major categories. This dual classifier performed better for taxa with low relative abundance than a single classifier fed with a unique feature vector containing stacked shape and texture information. Although this work employed a different approach for combining classifiers' results than the "classification by committee" technique of Ellis et al. (1997), they both documented that when using classifiers in parallel and merging their results with any strategy, a broader spectrum of image details can be covered, thus resulting in higher accuracies.

4.4 Classification via support vector machines

Support Vector Machines (SVMs) were introduced in the 1990s. The aim of an SVM is to find the hyperplane that classifies the given data points and has the maximum margin (the maximum distance between the data points of both classes). Contrary to LDA, an SVM makes no assumptions about the data at all. Moreover, if LDA produces an analytical solution, SVM-based classification is an optimization problem. An SVM does not make use of the entire input dataset. The subsets of data points that are closest to the hyperplane are called support vectors, and the SVM is optimized over these vectors. Then, the support vectors determine how the SVM discriminates between classes.

Luo et al. (2003) published the first work applying an SVM to classify the massive quantity of images collected with the Shadow Image Particle Profiling Evaluation Recorder (SIPPER) (Samson et al. (2001)). After performing noise suppression, they calculated general and specific descriptors to produce 29 features. Due to the binary nature of the images, they lacked texture information. They classified two datasets comprising six plankton species with a soft-margin SVM. The first was classified with 90% accuracy, and the second was classified with 75.1% accuracy. The low accuracy achieved on the second dataset was due to insufficient image quality and unknown objects that could not even be manually identified. To speed up the feature calculation process without compromising accuracy, they published an expansion of this work (Luo et al. 2004a) where they described a selection process to reduce the number of features from 29 to 15. Hence, they obtained similar accuracies for both datasets attaining a processing time of 160 s for 6000 images

with a one-versus-one approach. They acknowledged the need to develop new features for the higher-resolution version of SIPPER (25mm p/pixel). It is remarkable how the authors wisely picked a soft-margin SVM to cope with the noisy nature of the dataset. A soft-margin SVM tries to maximize accuracy while allowing some misclassification.

Hu and Davis (2005) employed co-occurrence matrix (COM) features with an SVM to classify images captured with the VPR. Co-occurrence matrices are textural features that describe the frequency at which one grey tone appears in a specified location in relation to another grey tone on an image (Haralick et al. 1973). Thus, their system computed eight COMs per image to perform posterior classification with an SVM. They obtained a 72% accuracy compared with the 61% accuracy of Davis et al. (2004). Both approaches classified the same dataset comprising seven categories, but the latter with an ANN was obtained from Tang and Stewart (1996). In another attempt, the authors combined texture features with shape-based features as classifier inputs, obtaining a mere 1% accuracy improvement over that achieved with COM features alone. After several tests, they concluded that the main accuracy improvement was achieved due to the COM features instead of the nature of the classifier, confirming that texture-based features are ideal for classifying field-collected images. This work highlighted that features should be selected depending on the given image's nature (captured in the laboratory or in situ); this step is more critical than selecting or tuning a specific classifier. Moreover, a COM has the advantage of low computational requirements. The COM-SVM technique reduced the manual identification process from weeks to several hours.

Lisin et al. (2005) tried using a new method to deal with plankton images in which occlusion and clutter were present. First, they computed global features based on shape and texture information and local features based on the scale invariant feature transform (SIFT). Then, the features were independently classified; the global features were classified by an SVM, and the local features were classified by a nonparametric density (NPD) classifier, attaining 55% and 52% accuracy, respectively. Later, the authors tested two methods for combining the previous classifiers' outputs: stacking and 2-tiered hierarchical classification. They obtained the highest accuracy (65.5%) with the stacking combination method on a 14-category dataset captured with the VPR (Fig. 7). Stacking consisted of concatenating the outputs of different classifiers (in this case, the SVM and NPD outputs) and using the result as an input vector (meta-features) for a meta-classifier (an SVM), which produced the final output. As in previous works (Ellis et al. 1997; Hu 2006), the

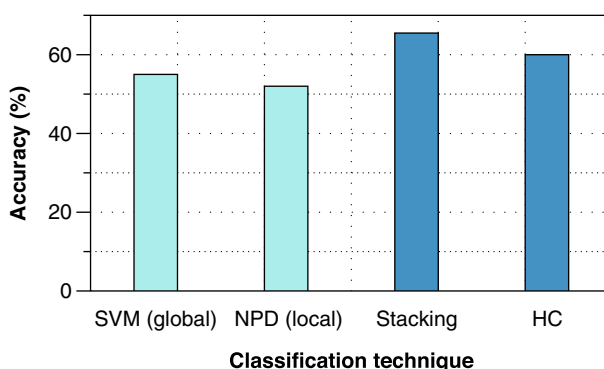


Fig. 7 Accuracies achieved by different classification techniques (Lisin et al. 2005)

Table 1 Summary of the reviewed papers

Paper	Pre-processing	Features	Classification technique	Processing power	Accuracy	Source of images	Groups Qty
Schlimpert et al. (1980)	–	Shape	LDA	Robotron 4200	93.2%	Vidicon	5
Jeffries et al. (1980)	Removal of appendages	Shape	LDA	–	93.1%	Vidicon	6
Jeffries et al. (1984)	–	Shape	LDA	6x MC68000 + Eclipse S/140	89%	Vidicon	8
Katsinis et al. (1984)	Removal of appendages	Shape	–	8x MC68000 + Central computer	92%	Vidicon	8
Chehdi et al. (1986)	Removal of appendages	Shape	HC	–	N/A%	Microscope	16
Chehdi and Coquin (1992)	Binarization	Shape	HC	–	78.8%	Microscope	14
Thonnat and Gandelin (1988)	Detection of the organisms	Shape	KBS	–	N/A%	Digitised photographs	14
Culverhouse et al. (1994)	Binarization, edition by hand	Shape	ANN	–	77%	Digitised photomicrographs	5
Ellis et al. (1994)	–	Shape, texture	ANN	–	73%	–	4
Culverhouse et al. (1996)	–	Shape, texture	RBF ANN	–	83%	Digitised photomicrographs	23
Tang and Stewart (1996); Tang et al. (1998)	Convolution, median filtering, out-of-focus objects rejection	Shape, texture	LVQ ANN	–	92.2%	VPR	6
Ellis et al. (1997)	Gabor filter	Shape, texture	Several ANN	–	70.3%	–	6
Dollfus and Beaufort (1999)	Translation, rotation, contrast normalization	–	Hierarchical ANN	Pentium 200 MHz	86%	–	13
Beaufort and Dollfus (2004)	Segmentation, translation, rotation	–	Parallel hierarchical ANN	–	91%	Microscope	11

Table 1 (continued)

Paper	Pre-processing	Features	Classification technique	Processing power	Accuracy	Source of images	Groups Qty
Hu and Davis (2006); Hu (2006)	–	Shape, texture	LVQ-ANN + SVM	–	64.7%	VPR	7
Luo et al. (2003, 2004a)	Denosing	Shape	SVM	Pentium 4 2.6GHz	75.6%	SIPPER	5
Hu and Davis (2005)	–	Texture	SVM	PC 1GHz	72%	VPR	7
Lisin et al. (2005)	Segmentation	Shape, texture, local	SVM	–	65.5%	VPR	14
Wang et al. (2006)	Lookup of interest points	Local	SVM	–	83%	–	5
Al-Barazanchi et al. (2015)	Resizing, data-augmentation	Generated by the CNN	SVM	–	96.7%	SIPPER	7
Cheng et al. (2020)	–	Generated by the CNN	SVM	–	98%	PlanktonScope	7
Luo et al. (2004b)	–	Shape	Assisted SVM	–	86.7%	SIPPER	5
Grosjean et al. (2004)	–	Shape, texture	Double bagging with LDA	Pentium 4 1.6GHz, 1GB RAM	74.6%	ZooScan	29
Bell and Hopcroft (2008)	–	–	RF	Pentium 4 3.2GHz, 4GB RAM	81.7%	Scanned	53
Gorsky et al. (2010)	ImageJ	–	RF	–	78%	ZooScan	20
Ye et al. (2011)	–	–	NBC	–	68%	ZooScan	32
Bochinski et al. (2019)	Segmentation for rotation only, data augmentation	–	CNN (based on AlexNet)	–	> 90%	ISHS	4
Dieleman et al. (2015)	Normalization, data augmentation	Shape, texture	ANN + CNN	Several Nvidia GPUs	Top-5 > 98%	ISHS	121
Dai et al. (2016)	Resizing, normalization, data-augmentation	–	11-layer CNN	–	93.7%	ZooScan	13

Table 1 (continued)

Paper	Pre-processing	Features	Classification technique	Processing power	Accuracy	Source of images	Groups Qty
Xiu and Zuoying (2016)	Rescaling, data-augmentation	-	32-layer Residual CNN	-	Top-5 95.8%	ISIIS	121
Dai et al. (2017)	Denoising, smoothing, Scharr operator, contrast optimization	Shape, texture	22-layer CNN (GoogLeNet)	Nvidia GTX 980Ti 6GB GPUs	96.3%	IFCB	30
Al-Barazanchi et al. (2018)	Resizing	-	18-layer CNN	-	98.2%	SIPPER	7
Cui et al. (2018)	Low and high pass filters, image enhancement	Shape, texture	8-layer CNN (AlexNet)	Nvidia Titan X Pascal GPU	96.6%	IFCB	30
Luo et al. (2018)	Contrast normalization, discard of noisy and low-confidence images, data augmentation	-	Spatially sparse 13-layer CNN	G2.2xlarge Amazon cloud (Nvidia GPU 1536 CUDA cores)	F1-score 0.881	ISIIS	37
Soh et al. (2018)	Data augmentation	-	19-layer CNN (based on YOLO)	Intel i7, Nvidia GTX980 GPU	IoU 86%	FlowCam, handmade	18
Lumini and Nanni (2019)	-	-	Fusion 8x CNN	Nvidia TitanX GPU	F-measure 0.953	IFCB, ZooScan, ISIIS	22, 20, 38
Li et al. (2021c)	Data-augmentation	-	CNN (based on YOLOv3)	Intel Xeon 5217, Nvidia RTX Titan GPUs	97.1%	IFCB	8

authors demonstrated that combining different classifier outputs resulted in higher accuracy. Nevertheless, they stated that stacking was rarely exploited in artificial vision applications. Although this ensemble technique delivered a low accuracy, it was something new in plankton classification.

Wang et al. (2006) also benefited from COM features, but this time they operated at the geometric and pixel levels instead of the texture level. First, they searched for the points of interest for use as the local extraction centres with a Harris detector. After performing normalization, they computed a 128-dimensional SIFT descriptor. Later, they generated patterns with an associative rule mining algorithm, keeping the most discriminant features by a Pearson correlation analysis. The authors called this local feature extraction process local co-occurring patterns (LCPs). Finally, they classified five phytoplankton species with an SVM-based one-versus-one approach, obtaining an accuracy of 83%. They expressed how this result was better than the 72% accuracy obtained by Culverhouse et al. (2003a, 2003b). Nevertheless, this comparison was inappropriate because the authors here categorized five species instead of the six species used in the cited work. It is interesting to note how during the heyday of SVMs, the authors reviewed thus far tried to achieve improved accuracy by researching new features instead of dealing with SVM parameters, which led to novel visual representations such as LCPs.

Unlike previous SVM frameworks, Al-Barazanchi et al. (2015) decided to substitute the time-consuming step of feature engineering with a CNN as a feature extractor. Hence, they fed a 10-layer CNN with resized images possessing 32x32 pixels, which were comprised of seven plankton categories captured with the SIPPER. Then, they trained this CNN until the maximum accuracy was achieved. From the obtained network, they employed each hidden layer as a feature for training another classifier. Thus, they separately evaluated a random forest and an SVM classifier with two different training sets, the original dataset and the same dataset extended via data augmentation (random image rotation), and with every layer as a feature vector. Finally, they obtained the best accuracy of 96.7% with the SVM classifier and the augmented training set using the features of hidden layer 1. With this technique, they outperformed other works' accuracies in terms of classifying the same dataset (Tang et al. 2006; Zhao et al. 2005, 2009, 2010; Li et al. 2014). This was the second reviewed paper, after the work of Dollfus and Beaufort (1999), that replaced the cumbersome feature extraction step with another original technique, allowing for the quick addition of new plankton species.

In a dataset, a plankton category contains dozens of images of the same species in multiple postures and positions. Nevertheless, a model requires even more images to attain a high classification accuracy. Some works, such as the previous study described above, employed data augmentation, which consists of artificially increasing the dataset's size by applying various transformations to the already labelled data, such as rotation, translation and flipping. Cheng et al. (2020) presented a new method to make a classifier invariant to plankton rotation, thereby eliminating the need for data augmentation. They first described each image with a polar representation, which consisted of depicting each pixel in terms of its distance and angle to the centre of the image, converting the rotational problem into a translational task. Thus, there exist classifiers such as CNNs that are robust to small translations due to their internal architectures (pooling layers). The combination of polar representations and their use in a CNN is a powerful technique for preventing the misclassification of two similar plankton images when one appears rotated relative to the other. With this idea in mind, the authors inputted each original image into a CNN and inputted its polar representation into another CNN. They trained each CNN independently and merged the outputs of their fully connected layers as the features for a subsequent classifier: an

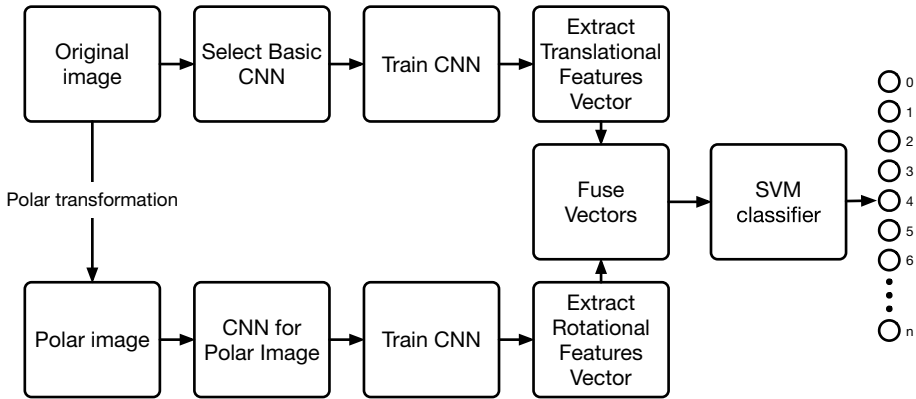


Fig. 8 Classification approach proposed by Cheng et al. (2020)

SVM (Fig. 8). Finally, the SVM obtained a 98% accuracy in terms of categorizing seven plankton species captured with PlanktonScope. This technique even outperformed a DenseNet CNN employing augmented data with 14 times the original size, which achieved 96.5% accuracy. This work solved one of the major issues of underwater artificial vision applications: the misclassification of similar objects captured at different angles. While polar representations were employed for plankton in the past (see the first reviewed paper, Schlimpert et al. (1980)), the authors presented an advancement based on the properties of CNNs. This framework can be valuable if combined with other techniques that address the problem of unbalanced categories in a dataset.

4.5 Semiautomated classification

Semiautomated classification aims to improve the accuracy of a classification model by combining the strengths of machines and human intelligence. Thus, a machine performs some tasks while a human performs other tasks, such as the initial grouping of data and the final classification decision. However, the requirement for human interaction is kept to a minimum. This approach is advantageous when labelled data are scarce, or new images are complex for a machine to label. This method's application dates to the early 1990s when researchers needed to perform simple classification tasks, such as sorting images into categories, with some input and guidance from human operators.

Luo et al. (2004b) updated their previous SVM classification approach (Luo et al. 2004a) by adding active learning. First, they built a multiclass SVM based on a training set and an unclassified set. Then, they ran the model, and an expert labelled the images from the unclassified set that were classified with the lowest confidence. Later, these manually labelled images were added to the training set and the model was retrained. This procedure was iterated several times until the desired accuracy of 86.7% was achieved. The main advantage of this method was that it required four times fewer images to produce the same accuracy as that of the classifier built on a random training set, as shown in Fig. 1 of their paper. Additionally, the iterative retraining process saved valuable time for the operators because it only requested labels for the most challenging images. This method is a reliable way to build a plankton dataset.

Grosjean et al. (2004) tested different automatic classifiers to categorize 29 groups of zooplankton. These images were acquired with ZooScan (Gorsky et al. 2009), a system built to scan biological samples. Among the 15 tested classifiers (such as LDA, HC, ANN), they obtained the highest accuracy of 74.6% with a double-bagging LDA classifier employing 27 textural and shape features. Nevertheless, the authors thought they should try another approach to attain higher accuracy due to species intravariability and the likelihood of training set mislabelling. Thus, they implemented a system to manually retag the items classified with low accuracy, improving the performance to 85% with a speed of 10,000 items in less than 5 s. The authors confirmed that human intervention is essential for managing the complexity of categorizing many species.

Bell and Hopcroft (2008) assessed the open-source ZooImage software on Alaska zooplankton to determine if it would be reliable for future research on salmon survival rates. ZooImage was based on the R project and the ImageJ software developed by the National Institutes of Health of the U.S. government to process biomedical images (Rasband 1997; Abramoff et al. 2004). First, they scanned the samples using a conventional flatbed desktop scanner (Epson Perfection 4990). Then, the authors used the default ZooImage configuration, which created vignettes of every particle (segmentation) found on the image for posterior feature extraction. In this study, no information about the type of features employed by ZooImage was included. However, they probably consisted of textural and shape features, as employed in other works using ZooImage (Table I, Fernandes et al. (2009)). Next, they created a training set and evaluated six classification algorithms: LDA, a recursive partitioning tree, K nearest neighbours, LVQ, an ANN, and a random forest (RF). They obtained the best accuracy of 81.7% with the RF algorithm by categorizing 53 species on field-collected samples, reducing the time required for quantitative analysis from hours to less than one hour. When they removed the discarded category, which accounted for more than 75% of all particles, the accuracy fell to 63.3%. Although they considered that the results were valuable, they anticipated that a semiautomated procedure was required to attain higher accuracies, expecting its inclusion in future software updates. As mentioned in this section's reviewed works, the semiautomated method is regarded as a technique for improving the accuracy of any model.

Gorsky et al. (2010) developed a new software for plankton classification: the Plankton Identifier (PkID). They performed a classification sequence composed of ZooScan, ZooProcess and Plankton ID. First, the images were acquired with ZooScan, as mentioned in a previous work. Next, they employed ZooProcess software (based on ImageJ software) to preprocess, segment, and extract features. Then, they manually created a training set with 30 categories, which was reduced to 20 due to the low accuracy shown on some categories in PkID cross-validation tests. Nevertheless, they kept some categories with low accuracy because of their ecological values. This reduction was made iteratively by merging categories with similar ecological values, resulting in higher accuracy (Fernandes et al. 2009). This training set achieved an overall accuracy of 78% using an RF algorithm. According to the authors, this performance was inaccurate for ecological studies because it is driven primarily by less abundant taxa. Thus, they decided to increase the number of categories from 14 to 42, assisting the machine with labelling when needed and resulting in 100% accuracy. The authors clearly showed how this semiautomated classification method is the best way to obtain the highest accuracy in taxonomic research at the expense of taking more time than the fully automated approach.

A year later, Ye et al. (2011) presented a semiautomated method based on a Bayesian probabilistic model. A Bayesian probabilistic model represents uncertainty about events using probabilities; it starts with a prior premise about an event (prior probability) and

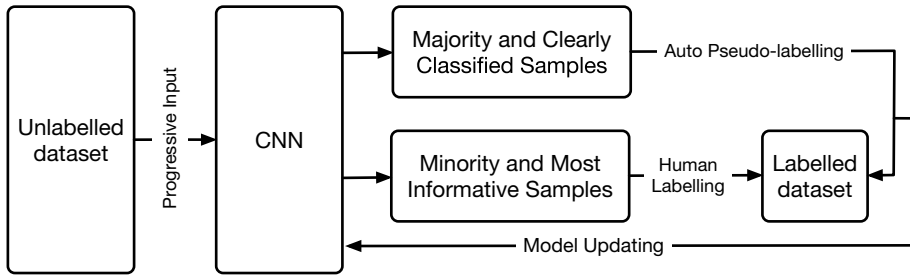


Fig. 9 Cost-effective active learning framework proposed by Wang et al. (2017)

then updates that premise with new information (likelihood) to obtain a new probability (posterior probability). After performing ZooScan image acquisition, the authors manually categorized every object, building unbalanced and balanced training sets comprised of 28 planktonic and nonliving categories. The balanced training set had the same number of images in each class. In contrast, the unbalanced class had different numbers of images in various classes. ZooProcess extracted features, and a naive Bayesian classifier was trained, obtaining a relationship between the posterior probability and cumulative recognition accuracy for each category. New images were automatically classified based on the posterior probabilities, obtaining accuracies of 69% and 68% for the unbalanced and balanced training sets, respectively. Second, over 31% of the images with the lowest confidence levels were manually reclassified by experts, obtaining 92% accuracy for both the unbalanced and balanced sets. This semiautomatic approach outperformed the method of Grosjean et al. (2004), and it offered the possibility of providing the results for any aggregation level with a single calculation step due to the Bayesian probability model. As stated in Bayesian probability model theory, this technique seems ideal for plankton classification because it can handle category intravariability, missing data in unbalanced categories, and noisy datasets. However, as embodied in the previously reviewed works, a semiautomated approach can raise any model to its maximum accuracy.

Bochinski et al. (2019) applied a variation of active learning to zooplankton for the first time: cost-effective active learning (CEAL, Fig. 9 (Wang et al. 2017)). Similarly, while active learning tries to maximize a model's accuracy regardless of the required human effort, CEAL tries to avoid unnecessary human intervention, hence the "cost-effective" aspect of the name. First, images were segmented using Otsu thresholding, but only when rotation was needed because the authors found that segmentation techniques removed plankton details. Then, they employed a CNN based on AlexNet and tested this method with two databases comprised of four zooplankton species acquired with the In Situ Ichthyoplankton Imaging System (ISIS) (Cowen and Guigand 2008). As a result, CEAL achieved the highest accuracy compared to the bare CNN and other active-learning methods; therefore, CEAL should be considered for future developments focused on semiautomated approaches.

4.6 Classification via convolutional neural networks

Convolutional Neural Networks (CNNs) are extensions of ANNs that can take an image as input. The name is derived from convolution, which is an operator that can merge two matrices by multiplying them. When correctly employed, convolutions can be used to

perform edge detection, image blurring or sharpening. The topology of a CNN considers multiple convolution layers to extract key image features. Interspersed with these layers, a CNN includes pooling layers, which are in charge of summarizing the features within a group of cells in the previous layer. Thus, the CNN also compresses the input data further. The final layer of the CNN serves as an input for an ANN to conduct pattern classification. CNNs were first introduced in the late 1980s by Yann LeCun and his colleagues at the Pierre and Marie Curie University in Paris and AT &T Bell Labs. They have become state-of-the-art approaches in computer vision applications, including image and video classification, object detection, and image generation.

Members of the reservoir laboratory (Ghent University - Belgium) won the US National Data Science Bowl (NDSB) competition by classifying 121 classes of plankton acquired with the ISIIS (Dieleman et al. 2015). First, they performed zero-mean unit-variance (ZMUV) normalization and data augmentation. ZMUV normalization is a common preprocessing step used in machine learning to ensure that the given data possess a consistent scale and have similar properties. The authors combined two different rescaling strategies due to the species' size diversity; one was based on image size, and the other was based on a fixed factor. For classification purposes, they employed a 13-layer CNN structure inspired by OxfordNet and an ANN employing the most discriminative conventional features, such as image sizes, image moments, and Haralick textures. Inspired by their previous participation in the Galaxy Challenge, they added cyclic pooling to the CNN to make the network more robust to the rotational plankton variations. Finally, they fused the results of these methods, obtaining a top-5 accuracy of over 98%, demonstrating their approach as a helpful tool for aiding experts in labeling tasks. They trained the models on Nvidia GPUs with convergence times from 24 to 48 h, depending on the model parameters. As a drawback, CNNs tend to overfit due to their complexity and the number of parameters relative to the training data, resulting in poor performance on new images. The authors overcame this issue for classes with as few as 20 samples per category by applying dropout, weight decay, data augmentation, pretraining, pseudolabelling, and parameter sharing techniques. In addition, cyclic pooling has become so powerful that neural networks centred on it have emerged, such as CyCNN and R2CNN. This work proved that high accuracy is possible with few samples, contrary to what is conveyed in deep learning theory.

Dai et al. (2016) presented ZooplanktoNet, a framework based on a CNN to classify a 13-class dataset acquired with ZooScan and provided by the Laboratoire d'Océanologie de Villefranche-sur-Mer, France. First, they normalized and augmented the dataset via rotation, translation, rescaling, shearing, and flipping. Then, they tested four popular CNN architectures on an augmented and nonaugmented dataset as reference points: AlexNet, CaffeNet, VGGNet and GoogleNet. The models' accuracies achieved on the augmented dataset were approximately 10% superior to those of the nonaugmented versions. Then, the authors chose AlexNet to continue their experiments but tested different optimizations for this CNN: different network depths, sizes and numbers of convolutional filters and the advantages of using local response normalization (LRN) and rectified linear units (ReLU). LRN is a preprocessing step that is commonly used in CNNs to prevent overfitting. A ReLU is an activation function that enables the learning of complex shapes and textures, improving overall model accuracy. The researchers obtained the best results with an 11-layer network, using LRN and parametric ReLU (PReLU) activation functions instead of the widely known ReLU function. The PReLU function resolves the dying ReLU problem and performs better, as shown in this work. Thus, this architecture was denoted as ZooplanktoNet, achieving an accuracy of 93.7% versus the 91.3% of the standard AlexNet

at the cost of requiring ten more minutes of training time (a total of 38 min). The authors documented how a standard CNN such as AlexNet could be adjusted to obtain higher accuracies depending on the nature of the given dataset. Nonetheless, this model was designed specifically for images acquired with ZooScan, with unknown performance for images acquired by other means.

Xiu and Zuoying (2016) applied the residual CNN (ResNet, He et al. (2016)) that was recently developed by Microsoft Research to the 121 classes of plankton dataset employed in the 2015 National Data Science Bowl (NDSB) competition. The idea of this work was to overcome the accuracy decrease induced when increasing the depth of a CNN to 20 layers or more. After rescaling and augmenting the dataset, they trained three residual CNNs with depths of 19, 32 and 50 layers plus a VGG-19 network as a reference, keeping the remaining parameters the same. The best network was the 32-layer network, with a top-5 accuracy of 95.8%. Although this accuracy was lower than the 98% achieved by the winning team in the NDSB competition, its computational cost was stunning, running 6.5x faster at a rate of 0.1 s per test set image. In general, ResNets perform better than traditional CNNs because they can train much deeper networks without incurring the vanishing gradient problem. The vanishing problem has been overcome by allowing the gradients to bypass one or more layers (as shown in Fig. 4 of their paper); this new internal structure is called “residual connection”. Moreover, during inference, ResNets are expected to run slower than traditional CNNs because they must perform calculations with both the original and residual inputs. They also use more layers, increasing the required computation time. Contrary to the above, the authors achieved a lower accuracy and a faster inference time compared to the VGG-19 network used as a reference. Perhaps this behaviour was due to the small dataset size, which the authors blamed for the result obtained in the 50-layer network test.

One year later, Dai et al. (2017) presented a classification system composed of three CNNs in parallel. The idea was to take advantage of textural features that describe differences when shapes do not and vice versa. To extract shape information, they first removed noise, smoothed shapes, applied the Scharr operator and optimized the contrast. To obtain texture information, they employed a Canny edge detector. Then, three CNNs were arranged in parallel; the first was fed with shape features, the second was

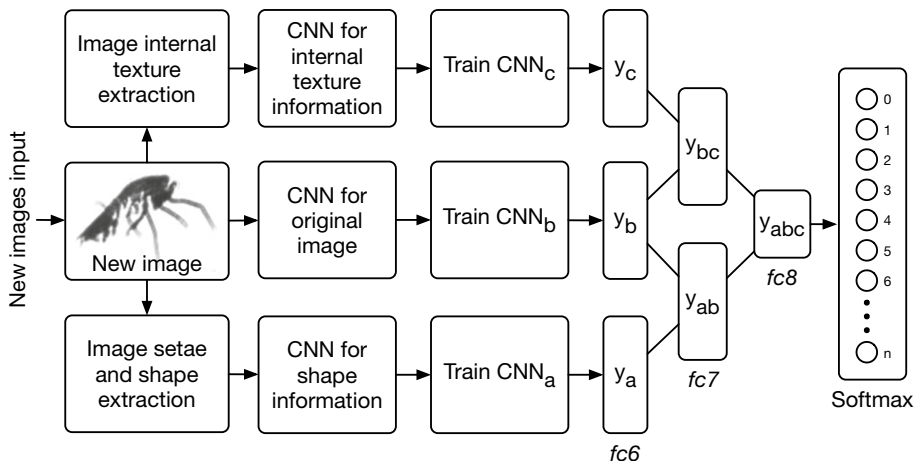


Fig. 10 Hybrid CNN based on AlexNet for plankton classification (Dai et al. 2017)

directly fed with the images, and the third was given textures, converging into a soft-max function (final layer) by a pyramid structure, as shown in Fig. 10. They classified a dataset containing 30 plankton classes acquired with the Imaging FlowCytobot (IFCB, Olson and Sosik (2007)). After completing several tests, they obtained maximum accuracies of 95.8% when employing networks based on AlexNet and 96.3% when using GoogleNet. AlexNet is an 8-layer CNN introduced in 2012 by researchers at the University of Toronto. It won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 with a top-5 error rate of 15.3%. GoogleNet is a 22-layer CNN introduced in 2014 by researchers at Google. It won the ILSVRC in 2014 with a top-5 error rate of 6.6%. All tests were run on four 6-GB Nvidia GTX 980Ti GPUs. Although the authors tried to employ as much image information as possible, they could barely achieve a 1% accuracy improvement over a single CNN (either AlexNet or GoogleNet) fed with only the original images as input, as shown in Tables 2 and 3 of the original paper. Thus, this method runs contrary to the trend of avoiding complicated feature calculations, showing no benefits.

Cui et al. (2018) presented an alternative approach to that used in the previous work. The authors also combined original images with shape and textural features as classifier information. However, they obtained shape features by applying Gaussian low-pass filtering and textural features by applying Gaussian high-pass filtering followed by logarithmic image enhancement. Then, they merged the shape features, the texture features, and the original image into a more complex feature representation using a concatenation layer. Finally, the authors input the merged vector into a single CNN based on AlexNet (Fig. 11). This approach obtained 96.6% accuracy, outperforming the framework based on AlexNet from Dai et al. (2017) by 0.8%. Although this procedure did not achieve a noticeable accuracy improvement, its simplicity compared to the previous method makes it preferable for future framework developments.

Al-Barazanchi et al. (2018) applied a CNN classifier to the same dataset employed in past works comprising seven plankton species (Tang et al. 2006; Zhao et al. 2005, 2009, 2010; Li et al. 2014; Al-Barazanchi et al. 2015). After resizing all the dataset images, they fed the images into an 18-layer CNN (based on VGGNet), achieving 98.2% accuracy on the testing dataset. Their CNN outperformed the best result from previous

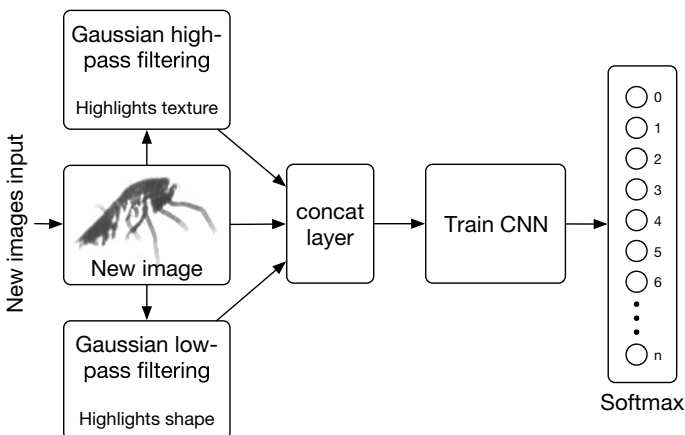


Fig. 11 Classification approach with a single CNN based on AlexNet (Cui et al. 2018)

works, which achieved 96.7% accuracy using a 10-layer CNN for feature extraction and an SVM as the classifier (Al-Barazanchi et al. 2015). This paper demonstrated how the unique use of a deep CNN outperforms any other method without data augmentation. However, it would have been interesting to compare this architecture with their previous work but instead using a CNN with the same depth as the feature extractor.

Luo et al. (2018) improved the parameters of the spatially sparse CNN formerly employed in the 2015 NDSB competition, which finished third in the final rankings. In a traditional CNN, each neuron in a layer receives input from all the neurons of the previous layer. In a spatially sparse CNN, each neuron in a layer receives input from a subset of the previous layer's neurons through a process called pruning. With these connection reductions, the network first learns the most important features. Then, it continues to learn more complex features with less important connections. Researchers created this type of CNN to recognize Chinese characters, where strokes were considered sparse pixels and the background was omitted for processing purposes. In this work, the authors considered a plankton image to be similar to a handwritten character, where most of it was background. In this way, a spatially sparse CNN could learn faster, with less overfitting, and with a lower computational cost than that of a dense CNN. They classified the 2015 NDSB competition dataset by regrouping the training set into 37 classes according to ecological significance. After conducting contrast normalization and discarding noisy and low-confidence images, data augmentation was performed on the rare classes via rotation, skewing, and scaling. Their 13-layer spatially sparse CNN obtained an F1-score of 0.881 and a precision of 90.7% on nonrare biological groups (23 classes), taking approximately 24 h for training and 165 h for prediction on new images using 1536 Nvidia CUDA cores. The result was quite impressive, considering the dataset size of 23.4 million images. However, this model may not be suitable for general plankton images, as they have high variability in their texture and illumination conditions, requiring more connections among neurons to process complex images. Moreover, this technique may soon become obsolete with the emergence of colour plankton databases, which present high degrees of variability.

Soh et al. (2018) presented a CNN architecture capable of identifying individual plankton in images with homogeneous clumping, heterogeneous interspersions or both. Homogeneous clumping occurs when plankton of the same species cluster together homogeneously. Heterogeneous interspersions occur when different plankton species are randomly distributed in a sample. These phenomena can affect the accuracy of plankton classification. Thus, the authors developed a 19-layer CNN based on the "You Only Look Once" (YOLO) architecture, which was capable of identifying and categorizing multiple objects in an image using just one computation. After conducting manual labelling, they performed data augmentation: six geometric distortions and two-pixel value modifications. Next, they applied this classifier to 18 species of plankton, detecting 100% of the validation set with an average intersection of union (IoU) of 86%. The IoU is a metric for object detection algorithms that measures the overlap between the predicted box and the ground-truth bounding box for an object in an image. This model took 20 h for training, running at 50 frames per second on the test set using one Nvidia GTX980. Due to the nature of the classifier, this method is a perfect fit for coping with scattered plankton, but distinguishing between clumped plankton and individuals is a different challenge, perhaps requiring extra preprocessing steps not covered in this work.

Lumini and Nanni (2019) tested a classification approach consisting of the fusion of different CNNs. They tested three plankton datasets on each of the following pretrained networks: AlexNet, GoogleNet, InceptionV3, VGG16, VGG19, ResNet50, ResNet101, and DenseNet. A pretrained network is a model that has already trained on a large dataset, the ImageNet

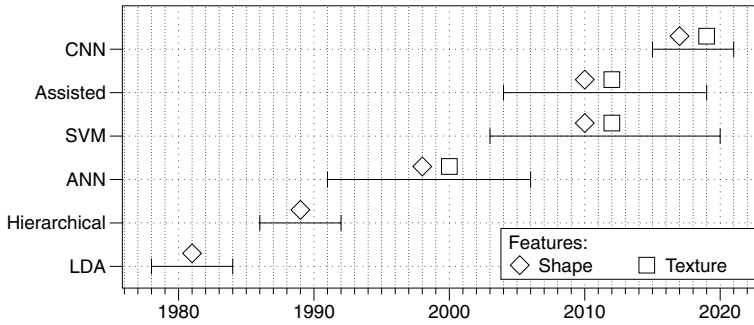


Fig. 12 Temporal coverage of reviewed papers

database in this case. Then, the parameters of these models were adapted to the new dataset. For this task, the authors employed the sequential floating forward selection (SFFS) feature picking method, which selects the most informative features from a pretrained models that are useful for a new dataset. The retraining step was performed under three different strategies: one round of tuning (1R), two rounds of tuning (2R) and preprocessing tuning (PR). 1R consisted of fine-tuning a pretrained model with a new dataset in a single round, 2R uses two rounds, and PR involves fine-tuning a preprocessed new dataset in a single round. For PR, the authors implemented preprocessing by applying data augmentation with four techniques: gradient, orient, local binary pattern, and local ternary pattern transformations. The three tested datasets were already used in past studies; they were acquired with the IFCB, ZooScan, and ISIIS. The best single pretrained model was DenseNet, but the highest accuracy was obtained with an ensemble method that consisted of employing the sum rule among the eight previous networks using one round and two rounds of fine-tuning, as shown in Table 6 of their paper (Fus_2R + Fus_1R). In a comparison, this approach obtained an F-measure of 0.953 versus the 0.9 achieved by Zheng et al. (2017) on the same IFCB dataset. They published another study that outperformed this ensemble's accuracy. It employed the SFFS feature selection approach but instead selected which classifier to add to the ensemble in each iteration based on its performance (Lumini et al. 2019). Thus, the SFFS ensemble consisted of 11 classifiers, obtaining an F-measure of 0.958 on the same IFCB dataset. As in the previously reviewed works, the authors revealed how an ensemble classifier could outperform individual CNNs. The use of pretrained models could be a powerful time-saving technique when the new target is a high-dimensional and complex dataset. Nevertheless, the complexity of running several networks to fuse their results later could be challenging. Further research may be needed to determine if the time required to implement this ensemble technique is shorter than that required to optimize a single CNN with layers adapted to the nature of the input images.

Li et al. (2021c) presented a new model that addressed two problems encountered in plankton classification: imbalanced categories and the loss of subtle features by a CNN during training. Imbalanced categories frequently occur in datasets for plankton classification tasks, where rare or poorly known taxa exist, negatively affecting a model's accuracy. Retaining subtle features, such as the shape of a head or the presence of appendages, is crucial for accurately classifying different plankton species. The authors solved the imbalance issue by creating fake images in the rare taxa categories using CycleGAN, a generative adversarial neural network that generates new images similar to those employed as inputs and described by Yann LeCun, Chief Artificial Intelligence Scientist at Facebook, as "the most interesting idea in the last ten

years in machine learning" (LeCun 2017). Then, the authors proposed a densely connected structure based on the YOLOv3 model to better capture the subtle features of plankton. In a densely connected structure, each layer is connected to every other layer, rather than just the previous and successive layers (as in traditional CNNs), allowing for a more robust representation of the input data. The authors applied this method to classify the IFCB dataset, obtaining an average precision of 97.1% for eight classes. This method outperformed others, with a detection time of 51 ms per testing image, which is suitable for real-time applications. Although the experiments were run in a desktop environment (Nvidia Titan RTX GPUs), they successfully integrated this method into a low-power Jetson Nano board, looking for its implementation in autonomous underwater vehicles (AUVs). This latest work marked a turning point in plankton image classification, presenting a model that not only offers high accuracy, but is also likely to be seen soon in an in situ device performing real-time plankton classification.

5 Conclusions and future work

Scientists began operating machines produced in other fields to quantify and classify plankton using pattern recognition techniques. The research started with Quantimet (IMANCO 1970), a system that was initially designed for analysing steel and then employed in multiple areas from materials to life sciences. Similar hardware but specifically made to digitize plankton specimens was developed after.

As Fig. 12 shows, early systems employed linear classifiers to group plankton by their shape features, but these techniques were still unable to achieve species-level identification due to poor image quality. In Europe, knowledge-based classifiers followed for plankton grouping, reaching a lower accuracy than linear discrimination but classifying twice the number of groups. In these studies, some authors removed appendages to ease the feature extraction process and separated the specimens connected by antennae or swimming legs. In the early nineties, ANNs were introduced to categorize plankton, achieving similar performance to that of taxonomists when combining shape features with texture features for the first time in field-collected specimens. Some authors focused on which features resulted in higher accuracy instead of fine-tuning their network parameters, and others combined the results of several ANNs running in parallel. SVMs were debuted in plankton classification during the same time period, using local features and CNN-like feature generators and outperforming past techniques. In addition to these approaches, researchers started to assist automatic classifiers with manual labelling for specimens with lower accuracy, achieving better results. Nonetheless, including a discarded category was a simple way to obtain these promising results (Hu and Davis 2005; Bell and Hopcroft 2008).

Recently, CNNs have defined a new era in image classification, shifting the focus of researchers from feature engineering to network parameter optimization. These networks run faster and obtain better results than those produced using a different classifier on the same dataset, as shown in the literature. The state-of-the-art accuracy achieved by a CNN for a 4-121 class problem exceeded 90%. Authors used existing architectures for plankton classification and produced new models adapted to this specific issue with parallel configurations or internal CNN structure modifications. Plankton recognition has lagged behind other applications, but Asian authors applied the YOLO architecture to plankton detection; this is the most popular object detection algorithm, which is able to process images in real time and with maximum accuracy compared to any other network developed to date.

Eventually, seventy years after Nishizawa et al. (1954) performed direct underwater observations, research groups became close to embedding high-end algorithms into underwater systems to categorize plankton in near real-time for the first time. Perhaps, the answer to complete plankton identification during this century will be a multipurpose system composed of acoustics, cameras, and DNA barcoding, where artificial vision would play a key role.

Future research on plankton image classification should focus on generating new datasets comprising high-quality colour plankton images, such as the DYB-PlanktonNet dataset. For this purpose, new tools and equipment will be necessary, such as the new in situ underwater imaging system developed by the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (Li et al. 2021a). Colour images offer more information to classifiers, making it easier to achieve higher accuracies. A second step should focus on integrating state-of-the-art classifiers into low-power, high-performance dedicated hardware for classifying plankton in situ and in real time. The above strategies, together with advanced energy harvesting modules, will enable the automatic monitoring of plankton worldwide at different depths under an international agenda, such as the Argo program (Scripps Institution of Oceanography 2000). The Argo program is an international project that collects ocean parameters between the bottom of the ocean and the ocean's surface through approximately 4000 autonomous robotic instruments. If such a device is developed, it would be a game changer in terms of studying and understanding plankton populations, potentially revolutionizing our understanding of the ocean and its role in the climate system.

Acknowledgements The authors wish to thank Alonso Hernández-Guerra for his firm support in the development of oceanographic technology. Special thanks to Laia Armengol for her help in the domain of plankton. This study has been funded by Feder of the UE through the RES-COAST Mac-Interreg project (MAC2/3.5b/314). We also acknowledge the European Union projects SUMMER (Grant Agreement 817806) and TRIATLAS (Grant Agreement 817578) from the Horizon 2020 Research and Innovation Programme and the Ministry of Science from the Spanish Government through the Project DESAFÍO (PID2020-118118RB-I00).

Author contributions DS-T wrote the main manuscript text. AB drafted the article and prepared figures. MG helped to draft the article and revised its content critically. SH made a substantial contribution to the introduction. All authors reviewed the manuscript and approved the final version.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declaration

Conflict of interest The authors have no conflict of interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with Image. *J. Biophoton Int* 11(7):36–42
- Al-Barazanchi HA, Verma A, Wang SX (2015) Performance evaluation of hybrid CNN for SIPPER plankton image classification. In: 2015 Third International Conference on Image Information Processing (ICIIP). IEEE, Warknaghat, pp 551–556. DOI: <https://doi.org/10/gjm6bq>
- Al-Barazanchi HA, Verma A, Wang SX (2018) Intelligent plankton image classification with deep learning. *Int J Comput Vis Robot* 8(6):561
- Battiti R, Colla AM (1994) Democracy in neural nets: Voting schemes for classification. *Neural Netw* 7(4):691–707. [https://doi.org/10.1016/0893-6080\(94\)90046-9](https://doi.org/10.1016/0893-6080(94)90046-9)
- Beamish P (1971) Quantitative measurements of acoustic scattering from zooplanktonic organisms. *Deep Sea Res Ocean Abstracts* 18(8):811–822. [https://doi.org/10.1016/0011-7471\(71\)90048-9](https://doi.org/10.1016/0011-7471(71)90048-9)
- Beaufort L, Dollfus D (2004) Automatic recognition of coccoliths by dynamical neural networks. *Marine Micropaleontol* 51(1–2):57–73
- Bell JL, Hopcroft RR (2008) Assessment of ZooImage as a tool for the classification of zooplankton. *J Plankton Res* 30(12):1351–1367
- Blackburn N, Hagstrom A, Wikner J et al (1998) Rapid determination of bacterial abundance, biovolume, morphology, and growth by neural network-based image analysis. *Appl Environ Microbiol* 64(9):3246–3255
- Blaschko M, Holness G, Mattar M, et al. (2005) Automatic In Situ Identification of Plankton. In: 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05) - Volume 1. IEEE, Breckenridge, CO, pp 79–86. DOI: <https://doi.org/10/fv8jvv>
- Bochinski E, Bacha G, Eiselein V et al (2019) Deep Active Learning for In Situ Plankton Classification. *Pattern Recognition and Information Forensics*. Springer, Cham, pp 5–15
- Buesseler KO, Antia AN, Chen M et al (2007) An assessment of the use of sediment traps for estimating upper ocean particle fluxes. *J Marine Res* 65(3):345–416. <https://doi.org/10.1357/002224007781567621>
- Chehdi K, Coquin D (1992) Pattern recognition by image analysis. Application to marine biology. In: Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems. IEEE Comput. Soc. Press, The Hague, Netherlands, pp 492–495. DOI: <https://doi.org/10/dff29b>
- Chehdi K, Boucher J, Hillion A (1986) Automatic classification of zooplankton by image analysis. In: ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol 11. Institute of Electrical and Electronics Engineers, Tokyo, Japan, pp 1477–1480. DOI: <https://doi.org/10/chmhqz>
- Cheng X, Ren Y, Cheng K et al (2020) Method for training convolutional neural networks for in situ plankton image recognition and classification based on the mechanisms of the human eye. *Sensors* 20(9):2592
- Cooke R, Terhune L, Ford J, et al. (1970) Technical report No. 172: An opto-electronic plankton sizer. Nanaimo, B.C. : Biological Station, Fisheries Research Board of Canada
- Cowen RK, Guigand CM (2008) In situ ichthyoplankton imaging system (I SIIS): system design and preliminary results: in situ ichthyoplankton imaging system. *Limnol Ocean: Methods* 6(2):126–132
- Cowen RK, Sponaugle S, Robinson KL, et al. (2015) PlanktonSet 1.0: Plankton imagery data collected from F.G. Walton Smith in Straits of Florida from 2014-06-03 to 2014-06-06 and used in the 2015 National Data Science Bowl (NCEI Accession 0127422). <https://doi.org/10.7289/v5d21vjd>
- Cui J, Wei B, Wang C et al (2018) Texture and shape information fusion of convolutional neural network for plankton image classification. In 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO). <https://doi.org/10.1109/OCEANSKOB.2018.8559156>
- Culverhouse P, Ellis R, Simpson R et al (1994) Automatic categorisation of five species of Cymatocylis (Protozoa, Tintinnida) by artificial neural network. *Marine Ecol Prog Series* 107:273–280
- Culverhouse P, Simpson R, Ellis R et al (1996) Automatic classification of field-collected dinoflagellates by artificial neural network. *Marine Ecol Prog Series* 139:281–287
- Culverhouse P, Williams R, Reguera B et al (2003) Do experts make mistakes? a comparison of human and machine identification of dinoflagellates. *Marine Ecol Prog Series* 247:17–25. <https://doi.org/10.3354/meps247017>
- Culverhouse P, Williams R, Reguera B, et al. (2003b) Expert and machine discrimination of marine flora: a comparison of recognition accuracy of field-collected phytoplankton. In: 2003 International Conference on Visual Information Engineering VIE 2003, pp 177–181. DOI: <https://doi.org/10/cgmsjh>
- Culverhouse P, Williams R, Benfield M et al (2006) Automatic image analysis of plankton: future perspectives. *Marine Ecol Prog Series* 312:297–309

- Culverhouse PF, Herry V, Reguera B, et al. (2000) Dinoflagellate categorization by artificial neural network (DiCANN). Proceedings of the EurOCEAN 2000 conference pp 239–240
- Dai J, Wang R, Zheng H et al (2016) ZooplanktoNet: Deep convolutional network for zooplankton classification. Oceans 2016 Shanghai. <https://doi.org/10.1109/OCEANSAP.2016.7485680>
- Dai J, Yu Z, Zheng H et al (2017) A Hybrid Convolutional Neural Network for Plankton Classification. In: Chen CS, Lu J, Ma KK (eds) Computer Vision - ACCV 2016 Workshops. Springer, Cham, pp 102–114
- Davis C, Gallager SM, Berman M et al (1992) The video plankton recorder (VPR): design and initial results. Arch Hydrobiol Beih Ergebn Limnol 36:67–81
- Davis C, Hu Q, Gallager S et al (2004) Real-time observation of taxa-specific plankton distributions: an optical sampling method. Marine Ecol Prog Series 284:77–96
- del Giorgio PA, Duarte CM (2002) Respiration in the open ocean. Nature 420(6914):379–384
- Dieleman S, van den Oord A, Korshunova I, et al. (2015) Classifying plankton with deep neural networks. <https://benanne.github.io/2015/03/17/plankton.html>
- Dollfus D, Beaufort L (1999) Fat neural network for recognition of position-normalised objects. Neural Netw 12(3):553–560
- Ellis R, Simpson R, Culverhouse P, et al. (1994) Expert visual classification and neural networks: can general solutions be found? In: Proceedings of OCEANS'94, vol 1. IEEE, Brest, France, pp I/330–I/334, DOI: <https://doi.org/10/bwqtkv>
- Ellis R, Simpson R, Culverhouse PF et al (1997) Committees, collectives and individuals: expert visual classification by neural network. Neural Comput Appl 5(2):99–105
- Falkowski PG (1994) The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth Res 39(3):235–258. <https://doi.org/10.1007/BF00014586>
- Fawell JK (1976) Electronic measuring devices in the sorting of marine zooplankton. Zooplankton fixation and elimination pp 201–206
- Fernandes JA, Irigoien X, Boyra G et al (2009) Optimizing the number of classes in automated zooplankton classification. J Plankton Res 31(1):19–29
- Folt CL, Burns CW (1999) Biological drivers of zooplankton patchiness. Trends Ecol Evol 14(8):300–305
- Fulton J (1972) Trials with an automated plankton counter. Fish Res Board Canada 29(7):1075–1078
- Garjjo JC, Hernández-León S (2015) The use of an image-based approach for the assessment of zooplankton physiological rates: a comparison with enzymatic methods. J Plankton Res 37(5):923–938. <https://doi.org/10.1093/plankt/fbv056>
- Goldberg E, Baker M, Denis LF (1952) Microfiltration in oceanographic research. J Marine Res 11(2):194–204
- Gorsky G, Picheral M, Grosjean P (2009) Optical scanning device for liquid biological samples, process of operation and computer program for a computer connected to said device. U.S. Patent US7555144B2
- Gorsky G, Ohman MD, Picheral M et al (2010) Digital zooplankton image analysis using the ZooScan integrated system. J Plankton Res 32(3):285–303. <https://doi.org/10.1093/plankt/fbp124>
- Grosjean P, Picheral M, Warembourg C et al (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. ICES J Marine Sci 61(4):518–525
- Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics SMC- 3(6):610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
- Hastings JW, Sweeney BM, Mullin MM (1962) Counting and sizing of unicellular marine organisms. Annals New York Acad Sci 99(2):280–289. <https://doi.org/10.1111/j.1749-6632.1962.tb45313.x>
- Hays G, Richardson A, Robinson C (2005) Climate change and marine plankton. Trends Ecol Evol 20(6):337–344
- He K, Zhang X, Ren S, et al. (2016) Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp 770–778, DOI: <https://doi.org/10/gdcfkn>
- Hernández-León S, Olivar MP, Fernández de Puellas ML et al (2019) Zooplankton and micronekton active flux across the tropical and subtropical atlantic ocean. Front Marine Sci 6:535. <https://doi.org/10.3389/fmars.2019.00535>
- Hernández-León S, Koppelman R, Fraile-Nuez E et al (2020) Large deep-sea zooplankton biomass mirrors primary production in the global ocean. Nat Commun 11(1):6048
- Hu Q (2006) Application of statistical learning theory to plankton image analysis. PhD thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Woods Hole
- Hu Q, Davis C (2005) Automatic plankton image recognition with co-occurrence matrices and Support Vector Machine. Marine Ecol Prog Series 295:21–31

- Hu Q, Davis C (2006) Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. *Marine Ecol Prog Series* 306:51–61
- IMANCO (1970) The Quantimet 720 Image Analysing Computer
- Ingram RA, Preston K (1970) Automatic analysis of blood cells. *Sci American* 223(5):72–82
- Jacobs RA, Jordan MI, Nowlan SJ et al (1991) Adaptive mixtures of local experts. *Neural Comput* 3(1):79–87. <https://doi.org/10.1162/neco.1991.3.1.79>
- Jeffries HP, Sherman K, Maurer R et al (1980) Computer-processing of zooplankton samples. *Estuarine Perspectives*. Elsevier, Amsterdam, pp 303–316
- Jeffries HP, Sherman K, Maurer R, et al. (1981) Processing of zooplankton samples by electronic image analysis. In: *Proceedings of Second Inter-Agency Workshop on In-situ Water-quality Sensing*, Biological Sensors, ed. by ED Kennedy. NOAA/National marine pollution program office, Rockville, Maryland, USA, pp 59–80
- Jeffries HP, Berman MS, Poularikas AD et al (1984) Automated sizing, counting and identification of zooplankton by pattern recognition. *Marine Biol* 78(3):329–334
- Jerlov NG (1953) Influence of suspended and dissolved matter on the transparency of sea water. *Tellus* 5(1):59–65
- Jerlov NG (1955) The particulate matter in the sea as determined by means of the tyndall meter. *Tellus* 7(2):218–225
- Katsinis C, Poularikas AD, Jeffries HP (1984) Image Processing And Pattern Recognition With Applications To Marine Biological Images. In: Tescher AG (ed) *Applications of Digital Image Processing VII*, San Diego, p 324, DOI: <https://doi.org/10/gj6tn3>
- Kocak D, da Vitoria Lobo N, Widder E (1999) Computer vision techniques for quantifying, tracking, and identifying bioluminescent plankton. *IEEE J Ocean Eng* 24(1):81–95
- Kramer K, Goldgof DB, Hall LO, et al. (2011) Increased classification accuracy and speedup through pairwise feature selection for support vector machines. In: *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE, Paris, France, pp 318–324, DOI: <https://doi.org/10/d7dbh9>
- LeCunn Y (2017) Quora Session with Yann LeCun. <https://quorasessionwithyannlecun.quora.com/>
- Li J, Chen L, Liu P et al (2021) Development of a buoy-borne underwater imaging system for in situ meso-plankton monitoring of coastal waters. *IEEE J Ocean Eng*. <https://doi.org/10.1109/JOE.2021.3106122>
- Li J, Yang Z, Chen T (2021b) DYB-PlanktonNet. <https://iee-dataport.org/documents/dyb-planktonnet>
- Li Q, Sun X, Dong J, et al. (2019) PMID2019. <https://github.com/ouc-ocean-group/PMID2019>
- Li Y, Guo J, Guo X et al (2021) Plankton detection with adversarial learning and a densely connected deep learning model for class imbalanced distribution. *J Marine Sci Eng* 9(6):636
- Li Z, Zhao F, Liu J et al (2014) Pairwise nonparametric discriminant analysis for binary plankton image recognition. *IEEE J Ocean Eng* 39(4):695–701
- Lisin D, Mattar M, Blaschko M, et al. (2005) Combining Local and Global Image Features for Object Class Recognition. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, vol 3. IEEE, San Diego, CA, USA, pp 47–47, DOI: <https://doi.org/10/c3854f>
- Liu S, Thonnat M, Berthod M (1994) Automatic classification of planktonic foraminifera by a knowledge-based system. In: *Proceedings of the Tenth Conference on Artificial Intelligence for Applications*. IEEE, San Antonio, TX, USA, pp 358–364, <https://doi.org/10.1109/CAIA.1994.323653>
- Loke R, du Buf J, Bayer M et al (2004) Diatom classification in ecological applications. *Pattern Recognit* 37(6):1283–1285
- Lombard F, Boss E, Waite AM et al (2019) Globally consistent quantitative observations of planktonic ecosystems. *Front Marine Sci* 6:196
- Lumini A, Nanni L (2019) Deep learning and transfer learning features for plankton classification. *Ecol Inf* 51:33–43. <https://doi.org/10.1016/j.ecoinf.2019.02.007>
- Lumini A, Nanni L, Maguolo G (2019) Deep learning for plankton and coral classification. *Appl Comput Inf*. <https://doi.org/10.1016/j.aci.2019.11.004>
- Luo JY, Irisson JO, Graham B et al (2018) Automated plankton image analysis using convolutional neural networks. *Limnol Ocean: Methods* 16(12):814–827
- Luo T, Kramer K, Samson S, et al. (2003) Learning to recognize plankton. In: *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, vol 1. IEEE, Washington, DC, USA, pp 888–893, DOI: <https://doi.org/10/bds87v>
- Luo T, Kramer K, Goldgof D et al (2004) Recognizing plankton images from the shadow image particle profiling evaluation recorder. *IEEE Trans on Syst* 34(4):1753–1762

- Luo T, Kramer K, Samson S, et al. (2004b) Active learning to recognize multiple types of plankton. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, Cambridge, UK, pp 478–481 Vol.3, DOI: <https://doi.org/10/fqzqj2>
- MacLeod N, Association S (eds) (2007) Automated taxon identification in systematics: theory, approaches and applications. No. 74 in The Systematics Association special volume series, CRC Press, Boca Raton
- Maddux WS, Kanwisher JW (1965) An in situ particle counter. *Limnol Ocean* 10:R162–R168
- Maloney TE, Donovan EJ, Robinson EL (1962) Determination of numbers and sizes of algal cells with an electronic particle counter. *Phycologia* 2(1):1–8
- Mattern CFT, Brackett FS, Olson BJ (1957) Determination of number and size of particles by electrical gating: blood cells. *J Appl Physiol* 10(1):56–70
- McClatchie S, Thorne RE, Grimes P et al (2000) Ground truth and target identification for fisheries acoustics. *Fish Res* 47(2–3):173–191. [https://doi.org/10.1016/S0165-7836\(00\)00168-5](https://doi.org/10.1016/S0165-7836(00)00168-5)
- McKinnon KM (2018) Flow cytometry: an overview. *Curr Protoc Immunol* 120(1):5
- Mullin MM (1965) Size fractionation of particulate organic carbon in the surface waters of the western Indian Ocean. *Limnol Ocean* 10(3):459–462
- Murphy LS, Haugen EM (1985) The distribution and abundance of phototrophic ultraplankton in the North Atlantic 1,2: phototrophic ultraplankton. *Limnol Ocean* 30(1):47–58
- Ng CL, Chen QQ, Chua JJ et al (2017) A multi-platform optical sensor for in vivo and in vitro algae classification. *Sensors* 17(4):912
- Nielsen ES (1952) The Use of Radio-active Carbon (C14) for Measuring Organic Production in the Sea. *ICES J Marine Sci* 18(2):117–140
- Nishizawa S, Fukuda M, Inoue N (1954) Photographic study of suspended matter and plankton in the sea. Research report of the Faculty of Fisheries, Hokkaido University p 6
- Scripps Institution of Oceanography USD (2000) Argo. <https://argo.ucsd.edu>
- Olney JE, Houde ED (1993) Evaluation of in situ silhouette photography in investigations of estuarine zooplankton and ichthyoplankton. *Bull Marine Sci* 52:28
- Olson RJ, Sosik HM (2007) A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot: in situ imaging of nano- and microplankton. *Limnol Ocean* 5(6):195–203
- Ortner PB, Hill LC, Edgerton HE (1981) In-situ silhouette photography of Gulf Stream zooplankton. *Deep Sea Res A* 28(12):1569–1576
- Parsons TR (1965) An automated technique for determining the growth rate of chain-forming phytoplankton. *Limnol Ocean* 10(4):598–602
- Rabinowitch EI (1945) Photosynthesis and related. Interscience Publishers, Inc., New York
- Rasband W (1997) Image J. US National Institutes of Health, Bethesda
- Rehnberg BG, Schultz DA, Raschke RL (1982) Limitations of electronic particle counting in reference to algal assays. *Water Pollut Control Federation* 54(2):181–186
- Riley GA (1944) The carbon metabolism and photosynthetic efficiency of the earth as a whole. *Am Sci* 32(2):129–134
- Riley GA (1963) Organic aggregates in seawater and the dynamics of their formation and utilization. *Limnol Ocean* 8(4):372–381
- Rodenacker K, Hense B, Jütting U et al (2006) Automatic analysis of aqueous specimens for phytoplankton structure recognition and population estimation. *Microscopy Res Tech* 69(9):708–720
- Samson S, Hopkins T, Rensen A et al (2001) A system for high-resolution zooplankton imaging. *IEEE J Ocean Eng* 26(4):671–676
- Schlimpert O, Uhlmann D, Schüller M et al (1980) Automated pattern recognition of phytoplankton - procedure and results. *Int Rev der gesamten Hydrobiol Hydrograph* 65(3):427–437
- Sheldon R, Parsons T (1967a) A practical manual on the use of the Coulter Counter in marine science, fisheries research board of canada edn. Coulter Electronics Sales Company - Canada
- Sheldon RW, Parsons TR (1967) A Continuous Size Spectrum for Particulate Matter in the Sea. *J Fish Res Board Canada* 24(5):909–915
- Sieracki ME, Webb LK (1991) The Application of Image Analysed Fluorescence Microscopy for Characterising Planktonic Bacteria and Protists. *Protozoa and Their Role in Marine Processes*. Springer, Heidelberg, pp 77–100
- Simpson R, Culverhouse P, Ellis R, et al. (1991) Classification of euceratium gran. in neural networks. In: [1991 Proceedings] IEEE Conference on Neural Networks for Ocean Engineering. IEEE, Washington, DC, USA, pp 223–229, DOI: <https://doi.org/10/cgh9dq>
- Simpson R, Williams R, Ellis R et al (1992) Biological pattern recognition by neural networks. *Marine Ecol Progress Series* 79:303–308

- Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. A Series of books in biology, W. H. Freeman, San Francisco
- Soh Y, Song J, Hae Y (2018) Multiple Plankton Detection and Recognition in Microscopic Images with Homogeneous Clumping and Heterogeneous Interspersion. *Journal of the Institute of Convergence Signal Processing* p 7
- Sosik HM (2015) WHOI-Plankton. <https://github.com/hsosik/WHOI-Plankton>
- Suthers IM, Rissik D (2019) PLANKTON: a guide to their ecology and monitoring for water quality. Place of publication not identified, 2nd edn. CRC Press, Boca Raton
- Tang X, Stewart W (1996) Plankton image classification using novel parallel-training learning vector quantization network. In: OCEANS 96 MTS/IEEE Conference Proceedings. The Coastal Ocean - Prospects for the 21st Century, vol 3. IEEE, Fort Lauderdale, FL, USA, pp 1227–1236, DOI: <https://doi.org/10/cg6tg5>
- Tang X, Stewart W, Vincent L et al (1998) Automatic Plankton Image Recognition. *Artificial Intelligence for Biology and Agriculture*. Springer, Dordrecht, pp 177–199
- Tang X, Lin F, Samson S et al (2006) Binary plankton image classification. *IEEE J Oceanic Eng* 31(3):728–735
- Thonnat M, Gandelin M (1988) An expert system for the automatic classification and description of zooplanktons from monocular images. In: [1988 Proceedings] 9th International Conference on Pattern Recognition. IEEE Comput. Soc. Press, Rome, Italy, pp 114–118, DOI: <https://doi.org/10/fbpc6f>
- Tiselius P (1998) An in situ video camera for plankton studies: design and preliminary observations. *Marine Ecol Progress Series* 164:293–299
- Tiwari S, Gallager S (2003) Optimizing multiscale texture invariants for the identification of bivalve larvae. In: Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), vol 2. IEEE, Barcelona, Spain, pp III–1061–4, DOI: <https://doi.org/10/fbbrpp>
- Uhlmann D, Schlimpelt O, Uhlmann W (1978) Automated phytoplankton analysis by a pattern recognition method. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 63(4):575–583
- Wang H, Miller P, Culverhouse P (2006) Discovering the Local Co-occurring Patterns in Visual Categorization. In: 2006 IEEE International Conference on Video and Signal Based Surveillance. IEEE, Sydney, Australia, pp 6–6, DOI: <https://doi.org/10/bf5dv9>
- Wang K, Zhang D, Li Y et al (2017) Cost-effective active learning for deep image classification. *IEEE Trans Circuits Syst Video Technol* 27(12):2591–2600. <https://doi.org/10.1109/TCSVT.2016.2589879>
- Williams R, McCall H, Pierce R et al (1994) Speciation of the tintinnid genus *Cymatocyclus* by morphometric analysis of the loricae. *Marine Ecol Progress Series* 107:263–272
- Wolpert DH (1992) Stacked generalization. *Neural Networks* 5(2):241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
- Xiu Li, Zuoying Cui (2016) Deep residual networks for plankton classification. In: OCEANS 2016 MTS/IEEE Monterey. IEEE, Monterey, CA, USA, pp 1–4, <https://doi.org/10.1109/OCEANS.2016.7761223>
- Ye L, Chang C, Hsieh C (2011) Bayesian model for semi-automated zooplankton classification with predictive confidence and rapid category aggregation. *Marine Ecol Progress Series* 441:185–196
- Zarauz L, Irigoien X (2008) Effects of Lugol's fixation on the size structure of natural nano-microplankton samples, analyzed by means of an automatic counting method. *J Plankton Res* 30(11):1297–1303
- Zhao F, Tang X, Lin F, et al. (2005) Binary plankton image classification using random subspace. In: IEEE International Conference on Image Processing 2005, pp 1–357, DOI: <https://doi.org/10/cf7mzf>
- Zhao F, Lin F, Seah HS Bagging based plankton image classification. In: 16th IEEE International Conference on Image Processing (ICIP). IEEE, Cairo, Egypt, pp 2081–2084, 2009
- Zhao F, Lin F, Seah HS (2010) Binary SIPPER plankton image classification using random subspace. *Neurocomputing* 73(10–12):1853–1860
- Zheng H, Wang R, Yu Z, et al (2017) Automatic plankton image classification combining multiple view features via multiple kernel learning. *BMC Bioinformatics*. 18: 570 <https://doi.org/10/gghp3f>
- Zhou H, Wang C, Wang R (2008) Biologically-Inspired Identification of Plankton Based on Hierarchical Shape Semantics Modeling. In: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering. IEEE, Shanghai, China, pp 2000–2003, DOI: <https://doi.org/10/ffs2gq>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.