Development of Deep Learning Models for Predicting the Effects of Exposure to Engineered Nanomaterials on Daphnia magna

Karatzas, Pantelis; Melagraki, Georgia; Bradford, Laura; Lynch, Iseult; Varsou, Dimitra-Danai; Afantitis, Antreas; Tsoumanis, Andreas; Doganis, Philip; Sarimveis, Haralambos

DOI:
10.1002/smll.202001080

NANO · MICRO small
www.small-journal.com

Development of Deep Learning Models for Predicting the Effects of Exposure to Engineered Nanomaterials on *Daphnia magna*

Pantelis Karatzas, Georgia Melagraki, Laura-Jayne A. Ellis, Iseult Lynch,* Dimitra-Danai Varsou, Antreas Afantitis, Andreas Tsoumanis, Philip Doganis, and Haralambos Sarimveis**

This study presents the results of applying deep learning methodologies within the ecotoxicology field, with the objective of training predictive models that can support hazard assessment and eventually the design of safer engineered nanomaterials (ENMs). A workflow applying two different deep learning architectures on microscopic images of *Daphnia magna* is proposed that can automatically detect possible malformations, such as effects on the length of the tail, and the overall size, and uncommon lipid concentrations and lipid deposit shapes, which are due to direct or parental exposure to ENMs. Next, classification models assign specific objects (heart, abdomen/claw) to classes that depend on lipid densities and compare the results with controls. The models are statistically validated in terms of their prediction accuracy on external *D. magna* images and illustrate that deep learning technologies can be useful in the nanoinformatics field, because they can automate time-consuming manual procedures, accelerate the investigation of adverse effects of ENMs, and facilitate the process of designing safer nanostructures. It may even be possible in the future to predict impacts on subsequent generations from images of parental exposure, reducing the time and cost involved in long-term reproductive toxicity assays over multiple generations.

1. Introduction

Nanotechnology has emerged at the forefront of science and technology due to its enormous potential to produce revolutionary advances in material science and in several fields of application, including microelectronics, energy storage units, smart therapeutics, optical detection systems, and many others. Design of safe engineered nanomaterials (ENMs) is perhaps the most important challenge in the field,[1] because due to their small size, ENMs may result in the modulation of pathways and mechanisms of toxic action that may endanger human health and the environment. Nanoinformatics approaches have gained popularity over the last few years as novel tools to address several challenges in nanotechnology[2] including design of safer ENMs, based on computational and data analysis methodologies, with the goal of reducing to the greatest possible extent the need for traditional hazard and risk assessment methodologies that are based on animal testing.[3] Machine learning has been used extensively in nanoinformatics to develop predictive models for toxicity- and ecotoxicity-related endpoints, employing various approaches such as read-across methods,[4–6] nano-quantitative structure–activity relationships (nanoQSAR[7–10]), QSAR-perturbation models,[11–13] and workflows predicting molecular initiating events and key events in adverse outcome pathways (AOPs).[14] Among the different types of descriptors used in predictive modeling approaches, image descriptors resulting from the analysis of electronic images of ENMs have been employed successfully.[15,16]

P. Karatzas, D.-D. Varsou, Dr. P. Doganis, Prof. H. Sarimveis
School of Chemical Engineering
National Technical University of Athens
Athens 15780, Greece
E-mail: hsarimv@central.ntua.gr

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smll.202001080.

DOI: 10.1002/smll.202001080

Dr. G. Melagraki, D.-D. Varsou, Dr. A. Afantitis, A. Tsoumanis
Nanoinformatics Department
NovaMechanics Ltd.
Nicosia 1065, Cyprus
E-mail: melagraki@novamechanics.com

L.-J. A. Ellis, Prof. I. Lynch
School of Geography
Earth and Environmental Sciences
University of Birmingham
Birmingham B15 2TT, UK
E-mail: I.Lynch@bham.a.uk

One of the most promising new areas in artificial intelligence (AI) and machine learning for building predictive models are the so-called deep learning technologies,[17] which are extensions of the traditional neural networks architectures, using more hidden layers and a larger variety of activation functions,[18] that is, functions that map the input to the output response of each neuron. Convolutional neural networks (CNNs) have shown state-of-the-art performance for image classification, segmentation, and object detection and tracking.[19] Extremely accurate deep learning models have been created in many disciplines using only electronic images as input information. Huge effort is especially focused on medical (e.g., histopathology)[20,21] and microscopy images[22] as well as on the detection of common objects or humans by identifying, for example, pedestrians, vehicles, or faces[23] and more recently on the detection and count of whales from satellite or aerial images to guide conservation actions.[24] Applications of deep learning methodologies in nanoinformatics are very rare. Güven and Oktay applied CNN to distinguish Fe_3O_4 ENMs from background[25] and in a follow-up study, Oktay and Gurses[26] applied multiple output CNNs (MO-CNN) to detect the locations of Fe_3O_4 ENMs in electronic images, to provide their boundaries, and to define their size and shape based on the segmentation output.

Our work builds upon one of the most extensive ecotoxicological datasets available, using state-of-the art AI methodologies to address the problem of identifying the effects of exposure to coated or uncoated TiO_2, Ag, or AgS ENMs under different experimental conditions on *Daphnia magna*, based solely on electronic images. The ENMs had different surface coatings, and were dispersed in artificial *Daphnia* medium or representative test waters and exposed to *Daphnia* immediately or following 6 months of aging in the various waters. The aim was to develop a complete and fully automated workflow, which needs only an electronic image of the *Daphnia* as input and predicts if the daphnids are damaged or not, the severity of damage and the types of malformations present. The workflow was built upon various object detection and classification deep learning technologies, which were compared in terms of accuracy, robustness, flexibility, and computational costs.

The object detection part of the workflow aimed at detecting, isolating, and classifying regions of daphnids where specific malformations occur, such as the eye and the tail. We applied the single shot multibox detector (SSD) MobileNet neural network,[27] which is a flexible, light, and quick architecture that can run on any device efficiently and also the heavier, but more robust region with CNNs (R-CNN).[28] SSD takes one shot for detecting multiple objects within an image compared to R-CNN, which requires two shots for detecting the object—one for generating region proposals, and one for detecting the object of each proposal. Different deep learning architectures, namely ImageNet[29] and residual networks,[30] were employed for classifying cropped images into categories defining the type and the severity of malformations. The results illustrate that the deep learning models can identify with accuracy the regions on the *Daphnia* that could be affected after exposure to ENMs.

The proposed workflow can be used to screen multiple images and provide accurate predictions in very short computational times. Contrary to alternative image analysis approaches, deep learning methods do not require the user to pre-define and compute image descriptors. Descriptors are automatically detected throughout the training procedure. Training algorithms can process big sets of images (in the order of thousands) in relatively short computational times. On the other hand, the availability of big datasets is a prerequisite for developing successful and accurate deep learning models and the algorithms must be executed in high performance computing environments, employing graphics processing units (GPU).

The four deep learning models presented here are available through a Docker container. The complete workflow has been implemented as a user-friendly web application that allows easy and fast predictions. The application is available through the NanoSolveIT cloud platform, developed in the context of the H2020 project NanoSolveIT.[31]

2. Methods and Strategies

D. magna are a freshwater zooplankton found in lakes and ponds throughout the world, in Europe, Africa, Asia, and America. *Daphnia*, a keystone species used in regulatory testing, is widely used in ecotoxicological studies mainly due to its tractability under laboratory conditions, short generation time, small body size, and parthenogenetic life cycle, as well as its transparent body that allows for light microscopy imaging.

An ecotoxicological study based on the *Daphnia* reproductive test was initiated to derive useful information on the effect of various ENMs on the daphnids in different experimental conditions. A set of more than 4000 light microscopy images of *Daphnia* exposed to pristine and 2 year old ENMs in different media and at different time points over multiple generations was produced, collected, and organized to deliver a unique dataset for further in silico exploitation. Following the exposure of the parental generations (F0), the offspring from the third brood were split into two groups, one group that was continuously exposed to the same ENM for the subsequent generations (F_{exp}) while the other group was exposed only to the relevant test water over the subsequent generations to assess the potential for recovery from parental exposure (F_{rec}).

The dataset included *Daphnia* images from four generations, namely F0, F1, F2, and F3. Different media were also included in the study: high hardness (HH) combo medium and class V river water, respectively. The HH combo medium represents an average hard water standard without any natural organic matter (NOM) and is commonly used for the culturing of *Daphnia* while class V water is an artificial representative river water medium that has high alkalinity and high NOM concentrations (4.6 mg L^{-1}) and is representative of typical of waters found in the southern UK, Poland, Greece, France, the Balearic countries, and the Iberian Peninsula.[32] **Figure 1** shows schematically the data included in the study.

The *Daphnia* images were taken at different time points starting from day 3 post-exposure to the ENMs to day 24 with 3 days of interval. This resulted in images for eight different time points during the daphnid life cycle (growth and reproduction phases) and up to five images were taken for the same time point. Moreover, different ENM concentrations were used –5, 10, 20, 100, or 5000 μg L^{-1}. The details of the different studies are summarized in **Table 1**.

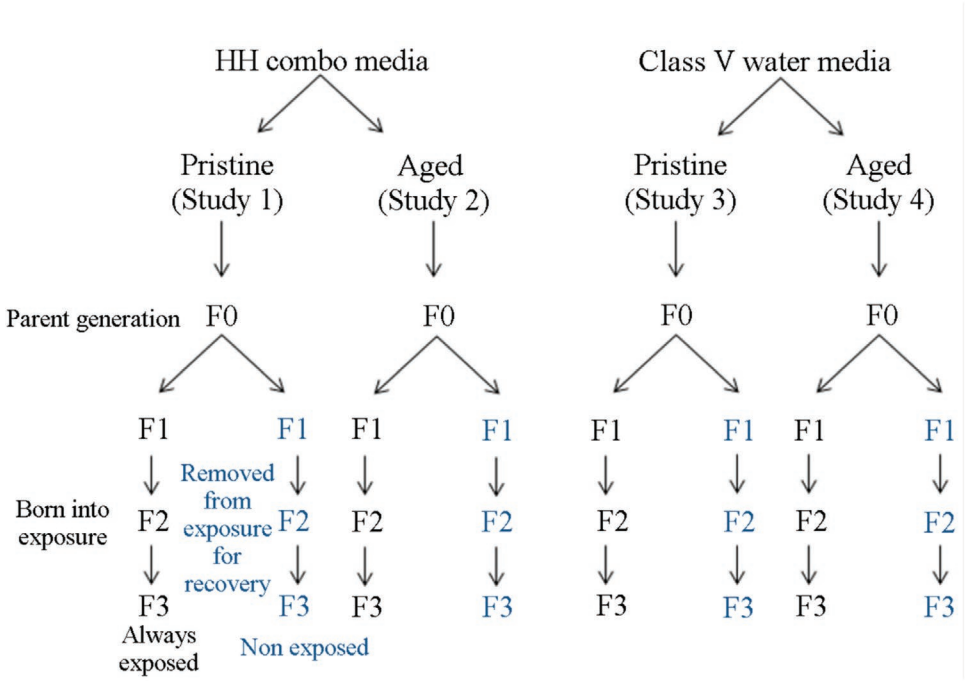

Figure 1. Data included in the different studies. In each case, the generation was formed from the third brood of the previous generation. In each study, the parent generation was exposed to ENMs and the F1 generation was split within 24 h of birth, with half continuing to be exposed to the ENM while the other half was further cultured in ENM-free medium to allow assessment of the potential for recovery from the parental exposure.

The ENMs used included three silver ENMs (uncoated Ag, Ag coated with PVP, and AgS) and two titanium dioxides (uncoated TiO_2, and TiO_2 coated with PVP), with polymeric PVP and micron-sized Ag used as coating-only and bulk (size) controls, respectively. All samples in each study were compared to the unexposed control. All ENMs were used in all three studies, while the polymer (PVP) and micron-sized bulk Ag controls were used only in studies 1 and 3.

More information regarding the life trait responses of the *Daphnia* to ENM exposure was collected from each experiment, including the total number of *Daphnia* in the first, second, third, and fourth broods; the time (in days) till the first, second, third, and fourth broods; and the average number of offspring per adult in each of the broods.

Data on the *Daphnia* images and the different physical malformations occurring in each of these images were organized in a format suitable for modeling (i.e., in an excel sheet with 1 row per treatment condition and time point) in an effort to in silico explore the available dataset and extract the maximum possible information from the available images. Changes in daphnid size relative to the control act as an indicator of toxicity, due to changes in feeding; when under stress daphnids may divert energy to antioxidant production to overcome oxidative stress and thus may not have enough energy to shed their carapace,[33] which is an essential step in growth (sheds every 3 days or so normally). Additionally, maternal feeding has been documented to affect offspring growth and reproduction.[34] Exposure to environmental toxicants has been linked with the activity of lipid allocation resulting in transcriptional and metabolic changes and enhanced lipid deposition.[35]

Table 1. Summarized details for each study.

Studies	Study details
Study 1	Pristine ENM, HH combo, four generations—all exposed
Study 2	Aged ENM, HH combo, four generations—only the first (F0) generation exposed
Study 3	Pristine ENM, class V, four generations—all exposed
Study 4	Aged ENM, class V, four generations—only the first (F0) generation exposed

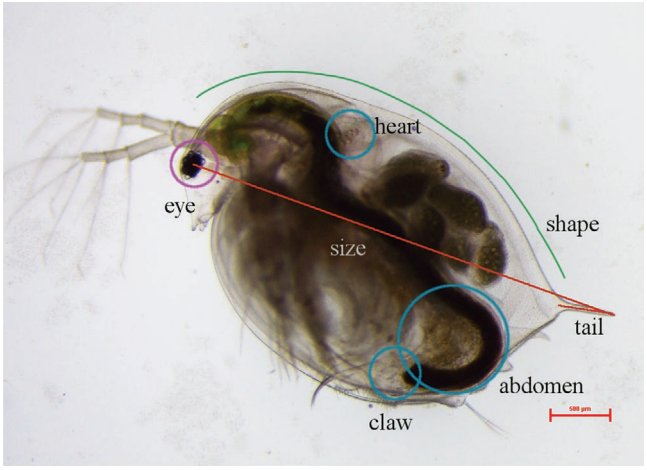

Figure 2. An example of a *Daphnia* image, where all parts of interest, that is, areas where malformations or changes as a result of ENM exposure can occur, are indicated.

Table 2. Types and classes of *Daphnia* malformations.

Area/region examined	Classification
Eye, shape	0 (normal) 1 (malformation)
Tail, overall size	0 (normal, if absolute value of (average control-measured value)/average control ≤10%) 1 (malformation, if absolute value of (average control-measured value)/average control >10%)
Abdomen, heart, claw	0 (normal) 1 (no lipids) 2 (less lipids than control) 3 (more lipids than control) 4 (more lipids than class 3)

An expert's opinion was first used to assess the type and severity of malformations captured by the light microscopy images. Malformations were observed to occur in the following regions of the daphnids: the eye (shape, presence/absence), the tail (length, presence/absence), the heart, the claw, and abdomen as well as changes in the overall daphnid shape and size (from head to tail) (**Figure 2**). Qualitative as well as quantitative assessment of the malformations that appeared in the images was proposed by expert evaluation (by the experimentalist who acquired the images), based on the comparison of images of *Daphnia* exposed to ENMs and control *Daphnia* (not exposed to ENMs), where the rest of the conditions (medium, time point, and generation) were exactly the same. For the eye, shape, heart, abdomen, and claws, the judgment on whether a daphnid was malformed or not was solely based on human expertise. The decision on the heart, abdomen, and claws was taken by visually comparing the number of lipids and grouping the images into four categories based on the malformation severity as determined by the expert, namely 1, if no lipids were detected; 2, if ENM-exposed *Daphnia* had less lipids compared to the unexposed control *Daphnia*; and 3 or 4, if ENM-exposed *Daphnia* had more lipids compared to the unexposed controls. For the eye and the overall size a visual inspection of the image classified the daphnid as normal or malformed. The length of the tail and the overall size were measured using ImageJ and compared with the averages over all the control *Daphnia* of the same category (i.e., same medium, time point, and generation). A relative difference greater than 10% indicated that the daphnid was malformed, otherwise the daphnid was normal. The types of malformations and the classes used by the expert are summarized in **Table 2**.

2.1. Workflow

The total dataset consists of 4323 images of *Daphnia*, of which 1219 images refer to controls or *Daphnia* that were not affected by the presence of ENMs, whereas 3104 images included *Daphnia* that have been affected due to the presence of ENMs. The *Daphnia* appearing in the images are not necessarily located at the center of the image and most of them have some background noise from salts or organic matter in the medium, as well as the ENM in some cases. Also the angle of the organism varies from image to image. That produces a relatively high-dimensional feature space and formulates a challenging object detection/image classification problem. To address this challenge, we designed a workflow that employs several deep learning architectures and methods and consists of several steps as shown in **Figure 3**.

The workflow starts with the development of an object detection deep learning model for drawing bounding boxes, isolating the objects of interest, and assigning the correct label to them. The objects of interest are the following: head, eye, heart, abdomen/claw, tail, tail-tip, and tail-base. We considered

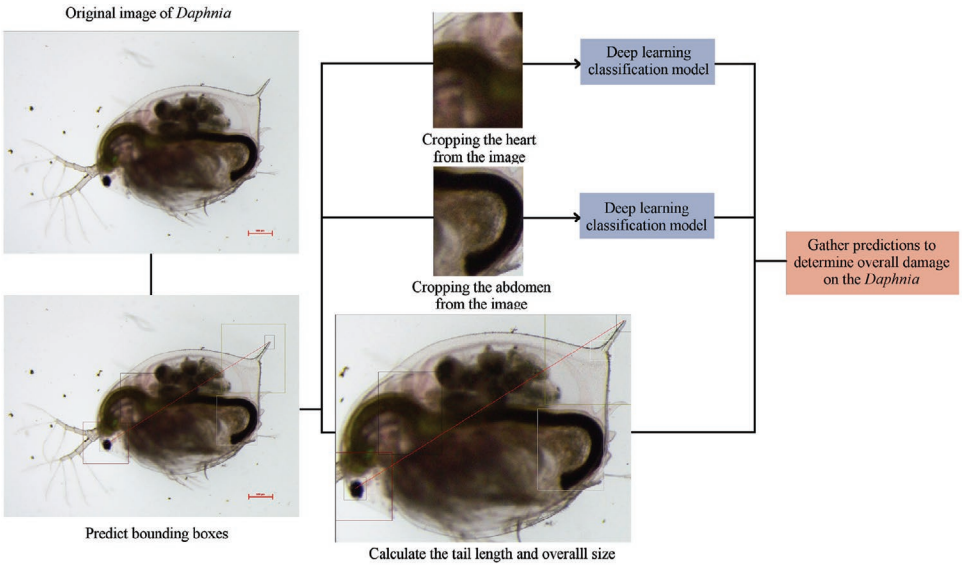

Figure 3. A graphical representation of the workflow used in the analysis. Emphasis was placed on extracting the maximum possible information from the image, by automating the malformation assessment and the decision-making procedure of the human expert. The workflow starts with detecting seven regions of interest in the image and continues with computing the overall size and the tail length. Cropped images of the abdomen/claw and heart regions are used as input to specially trained deep learning models to classify them with respect to lipid concentrations. The results are compared to control images to determine the types and levels of malformation.

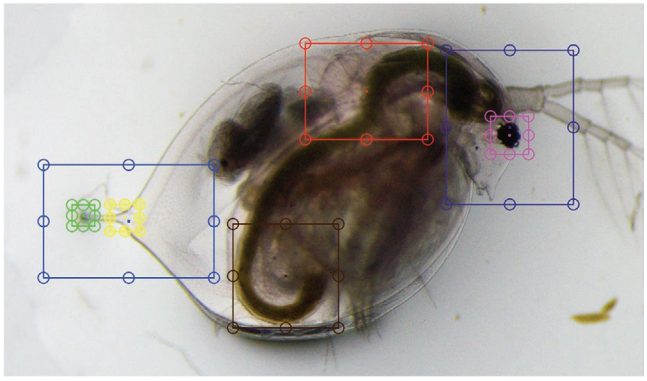

Figure 4. *Daphnia* and the annotated bounding boxes for the different features of interest used for classification of malformations in response to ENM exposure (head: purple box, eye: pink box, heart: red box, abdomen/claw: brown box, tail: blue box, tail-tip: green box, tail-base: yellow box).

the abdomen and the claw of the *Daphnia* as one object, since they are next to each other in the *Daphnia* body. For training this model, a dataset was constructed first, containing 518 randomly selected *Daphnia* images, the seven regions of interest for each image, and the respective annotations, using the standard PASCAL Visual Object Class (VOC) format. Pascal VOC format[36] is a standard format in the object detection domain that uses the extensible markup language (XML) for representing the information and has been adopted by many open source deep learning software packages, including the TensorFlow Object Detection API that was employed in this work. Each one of the images was manually segmented with respect to the seven object classes specified, and shown schematically in **Figure 4**, which presents an example of a *Daphnia* image with the annotated bounding boxes and their centers. The respective PASCAL VOC XML file contains the coordinates of the bounding boxes corresponding to the seven regions of interest and the respective annotations (the part of the body corresponding to each bounding box). Figure S1, Supporting Information, presents an example of a *Daphnia* image and parts of the PASCAL VOC XML corresponding to each region of interest.

Besides the boundaries that define each box, its center gives additional valuable information, because the distance between two boxes can be estimated by measuring the number of pixels from the center of one box to another and dividing by the appropriate distance scale. In particular, the distance between the tail tip and the tail base provides an estimation of the length of the tail, and the distance between the eye and the tail tip estimates the overall size of the specific daphnid. These measurements can be directly compared with the respective values of the healthy control *Daphnia*. Large discrepancies give indications that the daphnids are malformed due to their exposure to ENMs.

In all the object detection training runs, transfer learning[37] was employed. Training a completely new deep neural network requires huge amounts of data and a lot of computational power. Transfer learning is a technique that allows researchers to train neural networks with relatively smaller datasets and affordable computational power. To this end, an existing deep neural network model was employed that has been pre-trained on the big MS COCO dataset[38] that contains more than 300 K images and 2.5 million labeled instances of 91 object types. Only the last fully connected layer of the network was trained with the available data, while the first convolutional layers maintained the pretrained weights, which hold more primitive information regarding shapes. For each detected object, the produced model returns an array of four numbers that define completely a bounding rectangle that surrounds its position. The top value represents the distance of the rectangle's top edge from the top of the image, in pixels. The left value represents the left edge's distance from the left of the input image. The other values represent the bottom and right edges in a similar manner.

The object detection model outputs were used to train deep neural network models for predicting damage in each region of interest. The data on eye malformations were extremely unbalanced in favor of the normal versus the damaged class, so it was not possible to develop a model for prediction of eye malformation. Similarly, there are various and very different shape malformations such as misshaped back, larger body outline, misshaped front, and no tail, with very few images belonging to each subclass, and not enough for training a deep learning classification problem. Considering that in most of the images, the severity of damage provided by the expert for the abdomen and the claw were similar, we decided to consider these two regions as one class. Therefore, our approach is to detect following four possible malformations of the *Daphnia*:

i. Length of the tail
ii. Overall size of the daphnid
iii. Malformation in abdomen or claw
iv. Malformation in heart

The object detection deep learning model gives enough information to measure the length of the tail and the overall size and to compare them with the tail and size of the unexposed control daphnids. For the two last malformations, additional deep learning classification models had to be developed. New training sets were created by cropping the parts of the images that enclose the abdomen/claw and heart regions, as shown in Figure 4, and having a human expert assigning each image to a class.

The following five classes were defined based on the density of lipids in the cropped abdomen/claw images (examples are shown in **Figure 5**):

Class 0: no lipids are observed
Class 1: some lipids are observed near the intestine only
Class 2: lipids are formed in the whole abdomen area
Class 3: high concentration of lipids in the abdomen/claw area
Class 4: very high concentration of lipids in the abdomen/claw area

The following three classes were defined based on the density and location of lipids in the cropped heart images (examples are shown in **Figure 6**):

Class 0: no lipids are observed
Class 1: lipids are formed and observed near the food string
Class 2: lipids are formed and observed below and beyond the food string

The cropped images showing the abdomen/claw or the heart were split 90:10 into the training and test sets. To ensure a

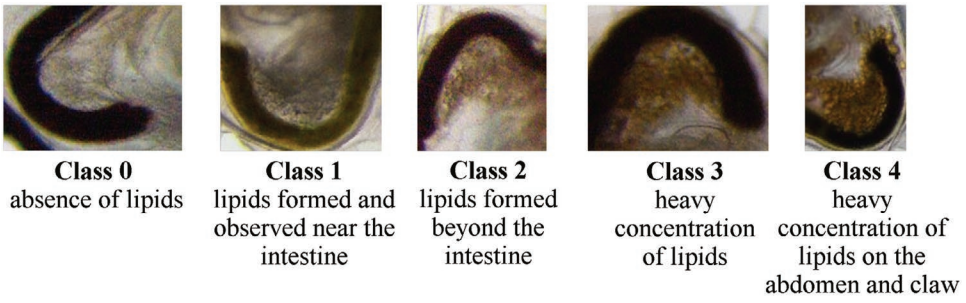

Figure 5. Examples of the five classes of lipids concentration in the abdomen/claw.

well-balanced representation of all classes in the training and test sets, 10% of the images from each class were selected randomly to populate the test sets and the remaining images constituted the training sets for the two classification problems. The residual networks and ImageNet spooling deep learning architectures were applied for training the classification models, which were validated on the test set of images.

3. Results and Discussion

The results from the deep learning models that were trained for the implementation of the workflow presented in the previous section, and the development of a completely automated tool, which is able to perform the object recognition and the malformation assessment procedure, are presented. All models were trained on a PC with an Intel Core i3 processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1070 GPU running on the GPU using the TensorFlow open source library for training deep learning machine learning models. Besides offering a large variety of deep learning architecture, TensorFlow has a number of additional advantages, for example, models can be containerized and run on GPUs or CPUs from a container equipped with TensorFlow serving.

3.1. Object Detection Deep Learning Model

For the object recognition problem, two deep learning architectures were employed, namely the single shot detector (SSD) mobile net v1[27] and Fast R-CNN.[39] The metric used for measuring the performance of the two algorithms was the total loss of each network, which is explained in more detail in the Experimental Section. Figures S2 and S3, Supporting Information, present the total loss of the two deep learning architectures as a function of training iterations. The model developed with the SSD architecture converged very quickly to a minimal loss function value (Figure S2, Supporting Information). In contrast, the Fast R-CNN diverged to very high values and was not able to recover after 8000 iterations (Figure S3, Supporting Information). Clearly, the SSD model, whose architecture is shown in **Figure 7**, was selected for object detection. The results are summarized in **Table 3**, which shows the number of images for which each region of interest was successfully detected. Figure S4, Supporting Information, presents examples of successfully annotated *Daphnia* images, using the developed workflow.

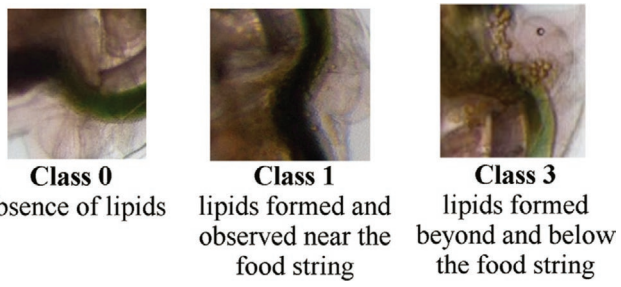

Figure 6. Examples of the three classes of lipids concentration in the heart.

3.2. Measuring the Overall Size of the *Daphnia* and the Length of the Tail

The size of an individual daphnid and the length of the tail are two critical points in the characterization of a daphnid as being damaged or not compared to the control measurements. The object detection model presented in the previous subsection provides all the information that allows automatic calculation of these critical parameters. More specifically, the overall size of a daphnid is calculated by counting the number of pixels between the centers of the eye and the tail tip bounding boxes and dividing by the appropriate scaling factor, as shown in **Figure 8**. In a similar manner, the length of the tail is computed by counting the number of pixels between the tail tip and the tail base. **Table 4** summarizes the results regarding the number of images for which the overall size and the tail length were automatically calculated successfully. Inability to calculate these characteristics was due to incorrect detection of objects, especially the tail-tip that is used in both tail length and overall size calculations (see Table 4), the appearance of more than one daphnid in a single image, the presence of background error, which sometimes is detected as an eye, or the absence of the scale line.

The accuracies of the automatic overall size and the tail length calculations are presented graphically as percentage error plots in **Figures 9** and **10**, respectively. The prediction errors were computed by comparing the modeling outcomes with the measurements of the human expert. The results show that both parameters can be predicted successfully, but better accuracy is obtained for the overall length estimation. This is due to the fact that tail length measurements are more sensitive to the location of the centers of the tail-tip and tail-base

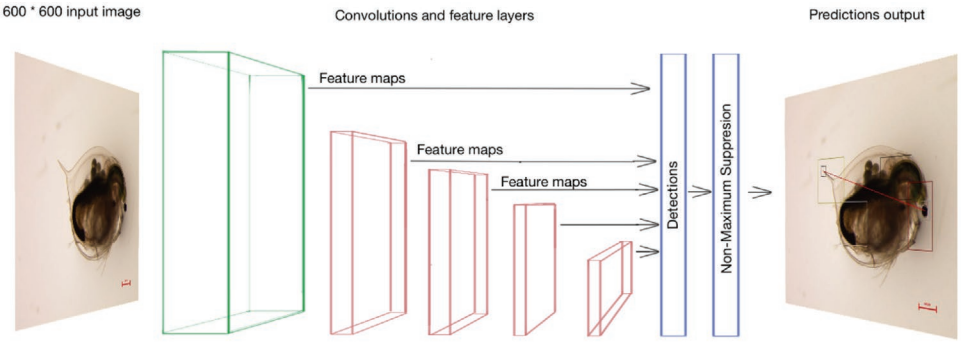

Figure 7. Illustration of the single shot detector (SSD) network architecture. The network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

bounding boxes as shown in Figures S5 and S6, Supporting Information. QQ plots of the overall size and tail length percentage errors are provided in **Figures 11** and **12** and illustrate that both of them follow normal distributions.

3.3. Classification of the *Daphnia* Abdomen/Claw or Heart in Terms of Lipids Concentration

The final step in the image analysis workflow is the development of deep learning models for the automatic classification of cropped *Daphnia* abdomen/claw or heart images according to the classes defined in the previous section, based on lipids presence and concentration. The models are developed using the same deep learning methods and only the models predicting malformation in abdomen/claw are presented in detail. The dataset consisted of 3780 cropped abdomen/claw images, after excluding 122 blurry images from the 3902 images, where the abdomen/claw region was detected correctly. The dataset was split into training and test sets containing 90% and 10% of the images, respectively, as explained in Section 2. Deep learning models were trained using the residual CNN algorithm and the ImageNet deep learning method and they were validated on the test set in terms of predicting the class related to lipid concentration, assigned by the human experts. The metric used for training the deep learning model was the overall accuracy, that is, the percentage of correct classifications. The two architectures had comparable results, with residual CNNs presenting a slightly better prediction accuracy (Figure S7, Supporting Information), while ImageNet CNN was faster in convergence but was overfitted in earlier iterations (Figure S8, Supporting Information).

The residual CNN was selected for further analysis because of the slightly better performance. The residual CNN architecture is presented in **Figure 13**. Figures S9 and S10, Supporting Information, compare the accuracy in the training and the test sets as a function of iterations in the first 12 000 iterations. The accuracy in the training set is constantly increasing, but in the test set, it reaches a plateau and after a number of iterations, it starts declining, meaning that the model development procedure is performing overtraining rather than continuing the learning procedure from the actual data.

The deep learning model produced at iteration 12 000 using residual CNN was selected as the final model for the five-class classification problem. The model was validated using a number of metrics: the confusion matrix, the overall accuracy, the Mathews correlation coefficient, the sensitivity, and specificity. The two last metrics are standard metrics for binary classification problems, but can be extended to multi-categorical classification problems, by using the one against all approach, where the sensitivity and specificity are computed for each class separately. For example, for class 0, true positives (TP) are all class 0 images classified as class 0, false negatives (FN) are all class 0 images not classified as class 0, true negatives (TN) are all non-class 0 images not classified as class 0, and false positives (FP) are all non-class 0 images classified as class 0. The statistical results on the test set are presented in **Table 5**.

The validation metrics in Table 5 indicate that the predictive performance of the model is not highly satisfactory. It should be noted, however, that the classification problem is an "ordinal" problem, wherein the classes are inherently ordered in terms of lipid concentration (from no lipid to high lipid concentration in the abdomen and claw), although we cannot define meaningful numeric differences between them.[40] Most validation metrics reported in **Table 6** underestimate the prediction accuracy, because they assume that every misclassification is considered equally costly. The confusion matrix gives a clearer and more complete

Table 3. Performance of the SSD deep learning model on *Daphnia* object detection, showing the number of successful detections of each object of interest (absolute number and percentage (%) of images available.

	Number of images	Successful abdomen/claw detections	Successful heart detections	Successful eye detections	Successful head detections	Successful tail detections	Successful tail-base detections	Successful tail-tip detections
Normal	1219	1087 (89.2%)	1141 (93.6%)	1171 (96.1%)	1133 (92.9%)	1187 (97.4%)	1029 (84.4%)	800 (65.6%)
Malformed	3104	2815 (90.7%)	2811 (90.5%)	2958 (95.3%)	2898 (93.4%)	3019 (97.3%)	2590 (83.4%)	2088 (67.3%)
Total	4323	3902 (90.3%)	3952 (91.4%)	4129 (95.5%)	4031 (93.2%)	4206 (97.3%)	3619 (83.7%)	2888 (66.8%)

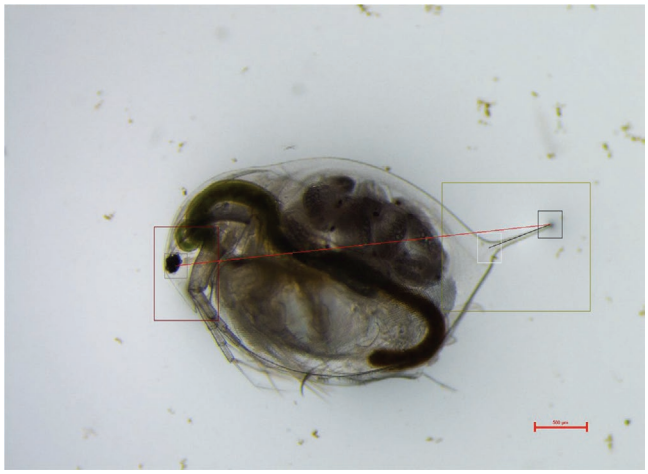

Figure 8. Measurement of the tail length and the overall size of the *Daphnia* based on object detection.

picture of the prediction accuracy of the model and indicates that besides the 46% of images that are classified correctly, another 41% of the predictions are next to the actual classes assigned by the expert, while only 13% of the predictions differ from the actual classes by more than one position in the ordered list.

We formulated an additional three-class classification problem by combining previously defined classes 1 and 2 into one class and classes 3 and 4 into another class, while class 0 remained the same (see Figure 5). The progress in the new deep learning model accuracy is shown in Figures S11 and S12, Supporting Information. The statistical results on the test set are presented in Table 6.

As expected, the accuracy has increased substantially and reached 75%.

Similar results were obtained after training deep learning models on the cropped heart images. The dataset consisted of 3930 images, after excluding 22 blurry images from the 3952 images, where the heart region was detected correctly. The dataset was split into training and test sets containing 90% and 10% of the images, as explained in Section 2. Figures S13 and S14, Supporting Information, present graphically the accuracy of the residual CNN network on the training and test sets. The model trained after 8000 iterations showed an overall accuracy of 67%. More detailed statistical results are shown in **Table 7**.

3.4. Implementation of the Deep Learning Workflow as a Service

All models and workflows presented in this work have been developed in the context of the H2020 Project NanoSolveIT.[31]

Table 4. Number of *Daphnia* images for which overall sizes and tail lengths were automatically calculated (absolute number and percentage (%) of total images available).

	Number of images	Number of images on which tail length and overall size were calculated successfully
Normal	1219	674 (55.3%)
Malfunctioned	3104	1738 (56.0%)
Total	4323	2412 (55.8%)

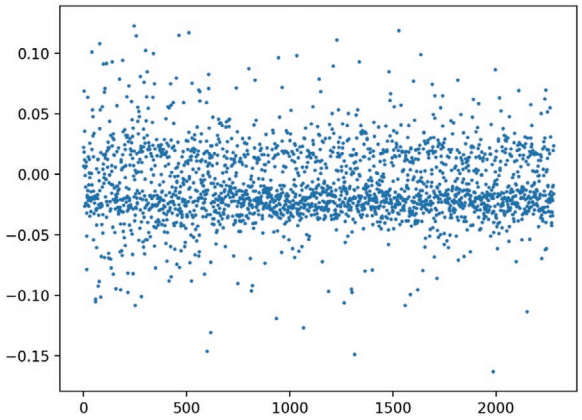

Figure 9. A plot of percentage errors for the overall *Daphnia* size measurements (expert determined (observed) values vs model outputs).

The SSD object detection model and the three residual CNN classification models (three-class and five-class abdomen/claw prediction models and the three-class heart prediction model) have been containerized in a Docker container available through https://hub.docker.com/r/nanosolveit/deepdaph-models. The complete workflow is being integrated into the NanoSolveIT cloud platform as a user-friendly web application, which is accessible through https://deepdaph.cloud.nanosolveit.eu/. The application is also available through application programming interfaces (APIs) that allow further integration with other services within the NanoSolveIT cloud platform but also with external services. The API documentation can be found at https://deepdaph-api.cloud.nanosolveit.eu/.

4. Conclusion

A multigeneration study on *D. magna* exposed to freshly dispersed and aged ENMs of various concentrations in different media was conducted. A series of the freshwater zooplankton *Daphnia* light microscopy images were produced and systematically studied in silico to predict the different malformations

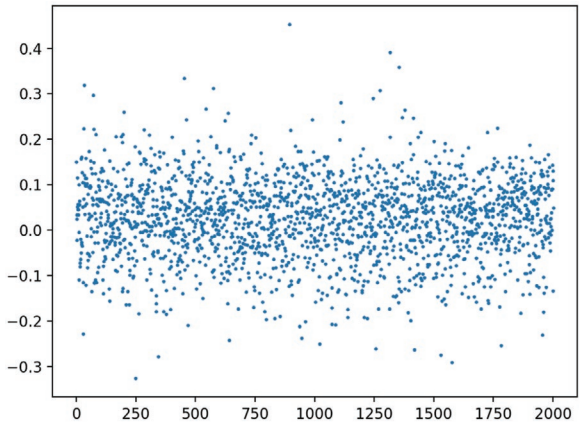

Figure 10. A plot of percentage errors for the tail length measurements (expert determined (observed) values vs model outputs).

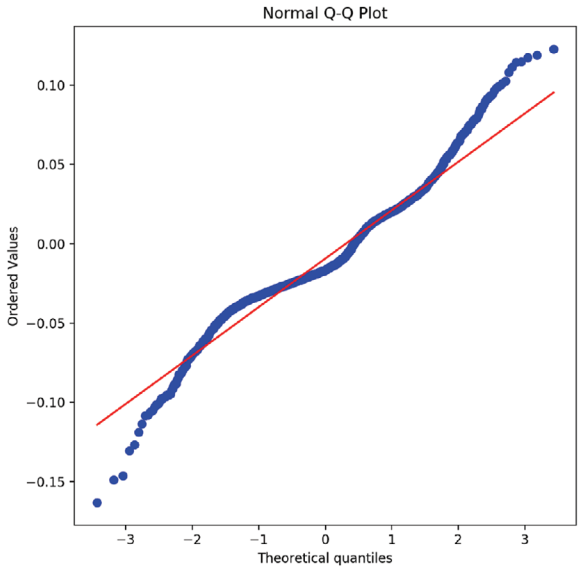

Figure 11. QQ plot of the percentage errors for the overall *Daphnia* size measurements.

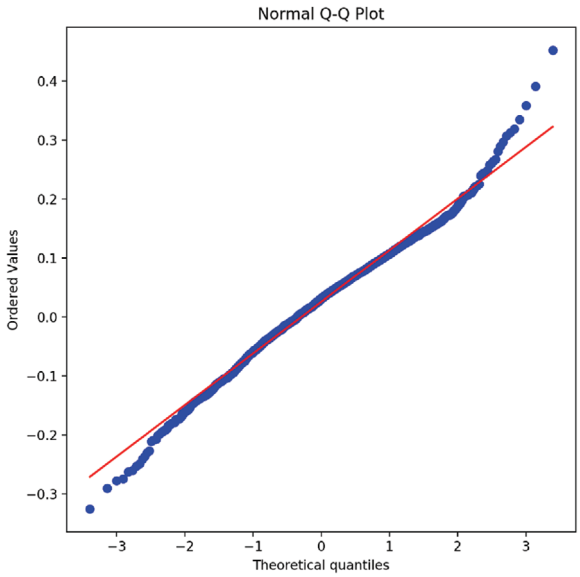

Figure 12. QQ plot of the percentage errors for the tail length measurements.

observed on the *Daphnia* including overall size; tail length; and malformations in heart, abdomen, and claw.

For this purpose, a computational workflow based on AI and deep learning methodologies has been developed to first detect, isolate, and classify regions of interest on the *Daphnia* images where specific malformation occurs and then to assess the type and the severity of malformations compared to the control.

To the best of our knowledge, this in silico study is the first attempt reported to develop validated deep learning predictive models to an extensive ecotoxicological study for predicting the effects of the direct or parental exposure to ENMs on *Daphnia*. The proposed models can automate time-consuming procedures needed for image classification by human experts and thus accelerate hazard assessment and facilitate the development of safe-by-design ENMs. Future extension of the work will allow adverse effects prediction for subsequent generations based on parental exposure images reducing the time and cost involved in long-term reproductive toxicity assays over multiple generations.

5. Experimental Section

Daphnia Maintenance and Culturing: *D. magna* are invertebrates, and as such are not subject to requirements for ethical approval for use in experimental research. Stocks of *D. magna* were maintained using pools of third brood Bham2 strain (genetically identical), which originated from the University of Reading[41] and the Water Research Centre (WRc), Medmenham, UK. *D. magna* were kept in a controlled environment in a 20 °C temperature with 12 h light and dark cycles and were cultured in a standard HH media (HH combo)[42] and an artificial natural water representative of a class V river lowland water[43] that was refreshed weekly to ensure healthy culture maintenance. *D. magna* cultures were fed *Chlorella vulgaris* algae daily, to total 0.5 mg carbon between days 0 and7 (750 µL) and 0.75 mg (1.5 mL) carbon from day 7. Third brood neonates were removed from exposure within 24 h of birth and used to set up the following generation.

ENMs Exposure: For the multigenerational studies (Figure 1), each ENM type (pristine/aged) was exposed to 10 daphnids per 250 mL in three replicates (total of 30 daphnids per exposure) to an F0 parent generation. Exposure concentrations were EC_5 concentration (TiO_2 ENMs) or EC_{30} concentration (Ag ENMs). The third broods (F1) from the F0 generation were split to produce a continuously paired exposure

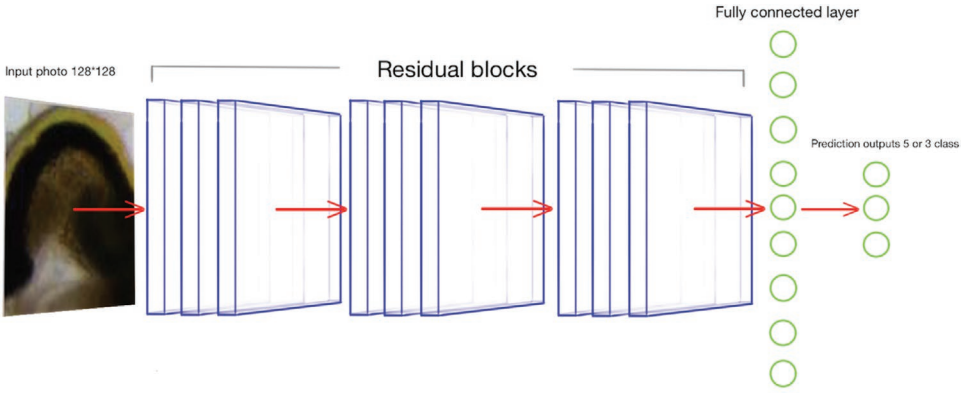

Figure 13. Schematic representation of the residual convolutional neural network (CNN) architecture using standard residual blocks. Each block is composed of two sets of batch normalization, ReLU, and convolutional layers. A fully connected layer completes the model.

Table 5. Validation metrics of residual CNN in the five-class classification problem for *Daphnia* abdomen/claw images.

Confusion matrix (absolute numbers and percentage (%) of test images available in each class, correct predictions are shown in bold).						
$n = 378$	Number of images	Predicted class 0	Predicted class 1	Predicted class 2	Predicted class 3	Predicted class 4
Actual class 0	80	**35 (43.8%)**	33 (41.2%)	10 (12.5%)	2 (2.5%)	0 (0%)
Actual class 1	74	7 (9.5%)	**37 (50.0%)**	19 (25.7%)	10 (13.5%)	1 (1.3%)
Actual class 2	99	3 (3.0%)	22 (22.2%)	**33 (33.3%)**	35 (35.4%)	6 (6.1%)
Actual class 3	83	3 (3.6%)	5 (6.0%)	7 (8.4%)	**43 (51.8%)**	25 (30.2%)
Actual class 4	42	0 (0%)	0 (0%)	3 (7.1%)	8 (19.1%)	**31 (73.8%)**
Five-class sensitivity and specificity						
	Class 1	Class 2	Class 3	Class 4	Class 5	
Sensitivity	43.8%	50.0%	33.3%	51.8%	73.8%	
Specificity	95.6%	80.3%	86.0%	81.4%	90.5%	
Accuracy: 0.46 and Matthews correlation coefficient: 0.3						

Table 6. Validation metrics of residual CNN in the three-class classification problem for *Daphnia* abdomen/claw images.

Confusion matrix (absolute numbers and percentage (%) of test images available in each class, correct predictions are shown in bold).				
$n = 378$	Number of images	Predicted class 0	Predicted class 1	Predicted class 2
Actual class 0	80	**70 (87.5%)**	8 (10.0%)	2 (2.5%)
Actual class 1	173	25 (14.4%)	**119 (68.8%)**	29 (16.8%)
Actual class 2	125	4 (3.2%)	25 (20.0%)	**96 (76.8%)**
Three-class sensitivity and specificity				
	Class 1	Class 2	Class 3	
Sensitivity	87.5%	68.8.0%	76.8%	
Specificity	90.3%	83.9%	87.7%	
Accuracy: 0.75 and Matthews correlation coefficient: 0.64				

Table 7. Validation metrics of residual CNN in the three-class classification problem for *Daphnia* heart images.

Confusion matrix (absolute numbers and percentage (%) of test images available in each class, correct predictions are shown in bold).				
$n = 393$	Number of images	Predicted class 0	Predicted class 1	Predicted class 2
Actual class 0	**148**	**89 (60.1%)**	57 (38.5%)	2 (1.4%)
Actual class 1	211	24 (11.4%)	**149 (70.6%)**	38 (18.0%)
Actual class 2	34	0 (0%)	6 (17.6%)	**28 (82.4%)**
Three-class sensitivity and specificity				
	Class 1	Class 2	Class 3	
Sensitivity	60.1%	70.6.0%	82.4%	
Specificity	90.2%	65.4%	88.9%	
Accuracy: 0.67 and Matthews correlation coefficient: 0.45				

(F_{exp}) over four successive generations (F0, $F1_{exp}$, $F2_{exp}$, and $F3_{exp}$) using the recovery ($F1_{rec}$) generation for three generations ($F1_{rec}$, $F2_{rec}$, and $F3_{rec}$). The media (with or without ENMs for the exposed and recovery experiments, respectively) was refreshed once a week.

Imaging of Daphnids for Assessment of Defects: Measurements of body size were taken every 3 days (between days 3 and 24) in accordance with moulting of the carapace.[44] Images were captured using a Nikon (Japan) stereomicroscope, model SMZ800 Digital Sight, fitted with a D5-Fi2 camera using NIS-Elements software. Scale bars represent 500 µm in all images. Five daphnids per treatment were imaged and body lengths were measured from the apex of the helmet to the base of the apical spine.

Image Analysis via Machine Learning: For the object detection problem, the fast R-CNN and SSD architectures were employed: The fast R-CNN algorithm is an improvement over the R-CNN method, proposed previously. R-CNN uses a selective search algorithm for extracting specific number (e.g., 2000) region proposals instead of selecting a very huge number of regions. The CNN acts as a feature extractor and the extracted features were fed into a Support Vector Machine model to classify the presence of the object within that candidate region. Fast R-CNN improves the convergence time by feeding the full image (and not the 2000 image proposals) to the CNN. This generates a convolutional feature map, which was used to identify the regions of

Table 8. The structure of the object detection deep neural network model.

Index	Name	Description
0	Locations	Multidimensional array of (7,4) floating point values between 0 and 1; the inner arrays representing bounding boxes in the form [top, left and bottom, right]
1	Classes	Array of seven integers (output as floating point values) each indicating the index of a class label from the labels file
2	Scores	Array of seven floating point values between 0 and 1 representing probability that a class was detected
3	Number of detections	Array of length 1 containing a floating point value expressing the total number of detection results

proposals, warp them into squares, reshape them to fixed sizes using a pooling layer, and finally feed them into a fully connected layer. A softmax layer predicts the classes of the proposed regions.

The SSD method discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. Compared to R-CNN methods, the SSD model is simpler, because it completely eliminates proposal generation and the subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. In the simulations, in this study, the SSD v1 pretrained model was used, employing the rectifier linear unit (ReLU) activation function and an initial learning rate of 0.004, which was decreased by a decay factor of 0.94 every 307 20 iterations.

The loss function used in both algorithms is a combination of confidence loss and localization loss. The localization loss between the predicted box and the ground truth box includes the offsets for the center points and the width and the height of the bounding box. The confidence loss is the softmax loss over multiple classes' confidences.

The produced object detection deep learning models take a *Daphnia* image as input. The configuration of the training method allows the input size of the image to be determined. The internal reshaper creates a 600×600 pixel image with three channels (red, blue, and green) per pixel. The reshaped image feeds the model as a flattened buffer of 270 000 byte values ($600 \times 600 \times 3$). Since the model is quantized, each value should be a single byte representing a value between 0 and 255. The output consists of the four arrays as shown in **Table 8**.

For the classification tasks, two main architectures were employed, namely the residual CNN and the ImageNet CNN. ImageNet neural networks are based on the LeNet architecture, which consists of stacking convolutional layers followed by pooling, ending with fully connected layers. ImageNet neural networks were the first to introduce ReLU as the activation function. Residual networks managed to go deeper by adding more layers and by introducing the residual layer in an effort to solve the problem of the vanishing gradient. Linear activation functions are also used in such architectures. The main difference here is the use of skip connections that form the residual block as the main component.

The produced ImageNet CNN had four convolutional layers each followed by a pooling layer and two fully connected layers with 2000 neurons each. The output of the network was a layer with dimensions equal to the number of classes. The ReLU activation function was used in all neurons of the neural network. The input tensor was either 512×512, 256×256, or 128×128. Each pooling layer was a 2×2 max pooling with strides of two. A variety of optimizers were applied ranging from simple gradient descent optimizers to more sophisticated approaches such as Adam or Adagrad. The learning rates started from 0.03 and decreased to 0.0003, but did not play a significant role on the predictive accuracy of the produced models. The main difference was noticed in the overall training time. For the more advanced optimizers, it was observed that the convergence time decreased at two-thirds of the time required by the simpler algorithms.

For the residual CNNs, standard residual blocks were used. The architecture starts with batch normalization using ReLU activation functions, which is followed by a convolutional block with a 3×3 kernel, normalizations and activation again, and then a convolutional layer again. Models with numbers of layers ranging from 10 to 20 with a stride every five layers were developed to downsample the input by the order of 2, similar to a pooling layer. Finally, a fully connected layer with 1000 neurons was added to complete the model. The main optimizer used for these models was an Adam optimizer with a learning rate of 0.001.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work received funding from the European Union's Horizon 2020 research and innovation programme via NanoSolveIT Project under grant agreement number 814572. The Daphnia microscopy images were generated via a NERC highlight topic grant (NE/N006569/1).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

deep learning, hazard assessment, image analysis, machine learning, nanoinformatics

Received: February 20, 2020
Revised: April 27, 2020
Published online:

[1] C. Micheletti, M. Roman, E. Tedesco, I. Olivato, F. Benetti, *J. Phys.: Conf. Ser.* **2017**, *838*, 012019.
[2] A. Afantitis, *Comb. Chem. High Throughput Screening* **2020**, *23*, 4.
[3] A. Haase, F. Klaessig, EU US Roadmap Nanoinformatics 2030, https://doi.org/10.5281/zenodo.1486012 (accessed: February 2020).
[4] N. Sizochenko, A. Mikolajczyk, K. Jagiello, T. Puzyn, J. Leszczynski, B. Rasulev, *Nanoscale* **2018**, *10*, 582.
[5] D. D. Varsou, A. Afantitis, A. Tsoumanis, G. Melagraki, H. Sarimvies, E. Valsami-Jones, I. Lynch, *Nanoscale Adv.* **2019**, *1*, 706.
[6] H. Wigger, B. Nowack, *Nanotoxicology* **2019**, *13*, 623.
[7] E. Burello, A. P. Worth, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **2011**, *3*, 298.
[8] A. Mikolajczyk, A. Gajewicz, B. Rasulev, N. Schaeublin, E. Maurer-Gardner, S. Hussain, J. Leszczynski, T. Puzyn, *Chem. Mater.* **2015**, *27*, 2400.
[9] G. Chen, M. G. Vijver, Y. Xiao, W. Peijnenburg, *Materials* **2017**, *10*, 1013.
[10] A. Afantitis, G. Melagraki, A. Tsoumanis, E. Valsami-Jones, I. Lynch, *Nanotoxicology* **2018**, *12*, 1148.
[11] V. V. Kleandrova, F. Luan, H. González-Díaz, J. M. Ruso, A. Melo, A. Speck-Planche, M. N. D. S Cordeiro, *Environ. Int.* **2014**, *73*, 288.
[12] V. V. Kleandrova, F. Luan, H. González-Díaz, J. M. Ruso, A. Speck-Planche, M. N. D. S Cordeiro, *Environ. Sci. Technol.* **2014**, *48*, 14686.

[13] R. Concu, V. V. Kleandrova, A. Speck-Planche, M. N. D. S Cordeiro, *Nanotoxicology* **2017**, *11*, 891.
[14] P. Nymark, L. Rieswijk, F. Ehrhart, N. Jeliazkova, G. Tsiliki, H. Sarimveis, C. T. Evelo, V. Hongisto, P. Kohonen, E. Willighagen, R. C. Grafström, *Toxicol. Sci.* **2018**, *162*, 264.
[15] K. Odziomek, D. Ushizima, P. Oberbek, K. J. Kurzydłowski, T. Puzyn, M. Haranczyk, *J. Microsc.* **2017**, *265*, 34.
[16] D. D. Varsou, A. Afantitis, A. Tsoumanis, A. Papadiamantis, E. Valsami-Jones, I. Lynch, G. Melagraki, *Small* **2020**, https://doi.org/10.1002/smll.201906588.
[17] Y. Lecun, Y. Bengio, G. Hinton, *Nature* **2015**, *521*, 436.
[18] K. Eckle, J. Schmidt-Hieber, *Neural Networks* **2019**, *110*, 232.
[19] H. Habibi Aghdam, E. Jahani Heravi, *Guide to Convolutional Neural Networks*, Springer, Basel, Switzerland **2017**.
[20] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, M. K. Khan, *J. Med. Syst.* **2018**, *42*, 226.
[21] F. Xing, Y. Xie, H. Su, F. Liu, L. Yang, *IEEE Trans. Neural Network Learn. Syst.* **2018**, *29*, 4550.
[22] R. Yamashita, M. Nishio, R. Kinh, G. Do, K. Togashi, *Insights Imaging* **2018**, *9*, 611.
[23] Y. Taigman, M. Yang, M. A. Ranzato, L. Wolf, presented at 27th IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, June **2014**.
[24] E. Guirado, S. Tabik, M. L. Rivas, D. Alcaraz-Segura, F. Herrera, *Sci. Rep.* **2019**, *9*, 14259.
[25] G. Güven, A. B. Oktay, presented at 26th Signal Processing and Communications Applications Conf. (SIU), Izmir, Turkey, May **2018**.
[26] A. B. Oktay, A. Gurses, *Micron* **2019**, *120*, 113.
[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, presented at European Conf. on Computer Vision, Amsterdam, The Netherlands, October **2016**.
[28] R. Girshick, J. Donahue, T. Darrell, J. Malik, presented at 27th IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, June **2014**.
[29] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Commun. ACM* **2017**, *60*, 84.
[30] K. He, X. Zhang, S. Ren, J. Sun, presented at 29th IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June–July **2016**.
[31] A. Afantitis, G. Melagraki, P. Isigonis, A. Tsoumanis, D. D. Varsou, E. Valsami-Jones, A. Papadiamantis, L.-J. A. Ellis, H. Sarimveis, P. Doganis, P. Karatzas, P. Tsiros, I. Liampa, V. Lobaskin, D. Greco, A. Serra, P. A. S. Kinaret, L. A. Saarimäki, R. Grafström, P. Kohonen, P. Nymark, E. Willighagen, T. Puzyn, A. Rybinska-Fryca, A. Lyubartsev, K. A. Jensen, J. G. Brandenburg, S. Lofts, C. Svendsen, S. Harrison, D. Maier, K. Tamm, J. Jänes, L. Sikk, M. Dusinska, E. Longhin, E. Rundén-Pran, E. Mariussen, N. E. Yamani, W. Unger, J. Radnik, A. Tropsha, Y. Cohen, J. Leszczynski, C. O. Hendren, M. Wiesner, D. Winkler, N. Suzuki, T. H. Yoon, J.-S. Choi, N. Sanabria, M. Gulumian, I. Lynch, *Comput. Struct. Biotechnol. J.* **2020**, *18*, 583.
[32] J. Hammes, J. A. Gallego-Urrea, M. Hassellöv, *Water Res.* **2013**, *47*, 5350.
[33] J. S. Garbutt, T. J. Little, *Biol. Lett.* **2014**, *10*, 20140356.
[34] F. Nasser, A. Davis, E. Valsami-Jones, I. Lynch, *Nanomaterials* **2016**, *6*, 222.
[35] N. Sengupta, P. D. Gerard, W. S. Baldwin, *Chemosphere* **2016**, *144*, 2302.
[36] M. Everingham, S. M. Ali Eslami, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, *Int. J. Comput. Vision* **2015**, *111*, 98.
[37] J. Talukdar, S. Gupta, P. S. Rajpura, R. S. Hegde, presented at 5th Int. Conf. on Signal Processing and Integrated Networks (SPIN), Noida, India, February **2018**.
[38] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, presented at Computer Vision—ECCV 2014, 13th European Conf., Zurich, Switzerland, September **2014**.
[39] S. Ren, K. He, R. Girshick, J. Sun, *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 6.
[40] J. S. Cardoso, R. Sousa, *Int. J. Pattern Recognit. Artif. Intell.* **2011**, *25*, 1173.
[41] L. H. Heckmann, R. Connon, T. H. Hutchinson, S. J. Maund, R. M. Sibly, A. Callaghan, *BMC Genomics* **2006**, *7*, 175.
[42] S. S. Kilham, D. A Kreeger, S. G. Lynn, C. E. Goulden, L. Herrera, *Hydrobiologia* **1998**, *377*, 147.
[43] J. Hammes, J. A. Gallego-Urrea, M. Hassellöv, *Water Res.* **2013**, *47*, 5350.
[44] D. Ebert, *Ecology, Epidemiology, and Evolution of Parasitism in Daphnia*, National Center for Biotechnology Information, Bethesda, MD **2005**.