Semi-supervised learning-based identification of the attachment between sludge and microparticles in wastewater treatment

Jia, Tianlong; Yu, Jing; Sun, Ao; Wu, Yipeng; Zhang, Shuo; Peng, Zhaoxu

DOI
10.1016/j.jenvman.2025.124268

Publication date
2025

Document Version
Final published version

Published in
Journal of Environmental Management

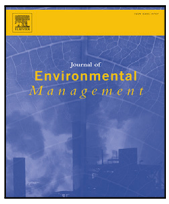
Citation (APA)
Jia, T., Yu, J., Sun, A., Wu, Y., Zhang, S., & Peng, Z. (2025). Semi-supervised learning-based identification of the attachment between sludge and microparticles in wastewater treatment. *Journal of Environmental Management*, *375*, Article 124268. https://doi.org/10.1016/j.jenvman.2025.124268

Journal of Environmental Management 375 (2025) 124268

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman



Research article

Semi-supervised learning-based identification of the attachment between sludge and microparticles in wastewater treatment

Tianlong Jia [a], Jing Yu [b], Ao Sun [c], Yipeng Wu [d,a], Shuo Zhang [a], Zhaoxu Peng [e,a,*]

[a] Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Water Management, Stevinweg 1, 2628 CN Delft, The Netherlands
[b] Erasmus University Medical Center, Department of Radiology and Nuclear Medicine, Dr Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
[c] Power China Zhongnan Engineering Corporation Limited, Changsha 410014, China
[d] School of Environment, Tsinghua University, 100084, Beijing, China
[e] Zhengzhou University, School of Water Conservancy and transportation, Kexue Road 100, Zhengzhou, 450001, China

ARTICLE INFO

Dataset link: https://github.com/TianlongJia/deep_pollutant_SSL, https://doi.org/10.5281/zenodo.13374998

Keywords:
Artificial intelligence
Computer vision
Self-supervised learning
Contrastive learning
Microparticles
Pollution

ABSTRACT

Monitoring the microparticle transfer process in wastewater treatment systems is crucial for improving treatment performance. Supervised deep learning methods show high performance to automatically detect particles, but they rely on vast amounts of labeled data for training. To overcome this issue, we proposed a semi-supervised learning (SSL) method based on the Simple framework for Contrastive Learning of visual Representations (SimCLR), to detect microparticles free from sludge and attached to sludge. First, we pre-trained a ResNet50 backbone by SimCLR, to extract features from much unlabeled data (1,000 images). Then, we constructed a Mask R-CNN architecture based on the pre-trained ResNet50, and fine-tuned it on a small quantity of labeled data (≈200 images with ≈600 annotated particles) in supervised learning fashion. We showcased its performance and practical applicability for microscopy images obtained from the water lab of TU Delft. The results demonstrate that the SSL methods obtain a significant improvement in mean average precision of up to 5% compared to the conventional supervised learning method, when a limited amount of labeled data is available (e.g., 91 labeled images). Furthermore, these methods improve the average precision for detecting attached particles by over 12%. With the detection results from the SSL methods, we measured the attachment efficiency of microparticles to sludge under varying mixed liquor suspended solids concentration and aeration intensity. The precise measurements demonstrate the effectiveness and practical applicability of the SSL method in facilitating long-term monitoring of particle transfer processes in biological wastewater treatment systems.

1. Introduction

Microplastics, plastic particles under 5 mm in size, are an emerging contaminant in water environment (Laskar and Kumar, 2019). They seriously harm aquatic organisms by causing suffocation and digestive issues, and may pose risks to human health through food chain transfer (Benson et al., 2022; Blackburn and Green, 2022). In the social water cycle, microplastics eventually enters the sewage treatment plant (STP) along the pipe network, which influences the treatment performance of STP (Hu and Chen, 2024; Li et al., 2023). Unlike the dissolved substrate which can be used directly, the particulate substrate usually has to be adsorbed by the sludge (floc, film or granule) firstly, then hydrolyzed into soluble substrate (Ortega et al., 2022), although sometimes the microorganisms even use microplastics as carbon source (Cydzik-Kwiatkowska et al., 2020). The adsorption of microplastics not only affects the wastewater treatment performance, but also determines the fate of microplastics in STPs. More than 90% of adsorbed microplastic fibers are retained in the activated sludge during dewatering since their size limits their transport (Keller et al., 2019), and the attached microparticles can also be stably retained in the sludge over the long term during anaerobic digestion (Feng et al., 2014). Thus, monitoring and controlling microplastics deserves attention in STPs to improve treatment performance (Alvim et al., 2020).

To monitor microparticles, researchers usually collected microscopy images and processed images using softwares, such as ImageJ (Campbell et al., 2019; Oliveira et al., 2018). However, employing processing software is time-consuming and labor-intensive, and requires specialized expertise for fine-tuning parameters manually (e.g., threshold, and output size). Recently, data-driven deep learning methods have recently

* Corresponding author at: Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Water Management, Stevinweg 1, 2628 CN Delft, The Netherlands.
E-mail address: pzx@zzu.edu.cn (Z. Peng).

https://doi.org/10.1016/j.jenvman.2025.124268
Received 13 October 2024; Received in revised form 16 December 2024; Accepted 19 January 2025
Available online 31 January 2025

gained significant attention for image processing and computer vision tasks, with studies highlighting their potential in wastewater treatment processes (Alvi et al., 2023). For example, Jia et al. (2024a) applied a Cascade Mask R-CNN model to detect and quantify microparticles free from biomass and entrapped in biomass from microscope images in biological wastewater treatment process. This model achieves a 13.8% improvement in micro-average precision, compared to the conventional ImageJ processing method. Additionally, the processing speed of the model is significantly faster than the ImageJ method with manual parameter tuning. Satoh et al. (2021) applied an Inception V3 model to automatically identify aggregated and dispersed flocs from microscopy images in full-scale STPs, achieving an accuracy of about 95%.

While the current results are encouraging, developing accurate deep learning models needs a substantial amount of labeled data for supervised learning. The expensive and labor-intensive manual labeling process has become a bottleneck in deep learning applications (Jia et al., 2023a). To address this issue, researchers are increasingly exploring the potential of self- and semi-supervised learning approaches, since they reduce dependency on the availability of a large amount of labeled data to obtain feature representations from data (Jing and Tian, 2020). Self-supervised learning extracts feature representations from images via well-designed pretext tasks without requiring manual labels (Misra and Maaten, 2020). The mainstream self-supervised learning approaches include two categories: generative and discriminative. The generative self-supervised approach learns feature representations by performing pixel-level reconstruction, but requires extensive computational resources (Goodfellow et al., 2014; Kingma and Welling, 2013). Discriminative self-supervised learning approaches learns feature representations using objective functions which are similar to those used in supervised learning method. This is achieved by employing pretext tasks, where both inputs and labels are derived from an unlabeled dataset (Chen et al., 2020).

Typical pretext tasks include relative position prediction (Doersch et al., 2015), Jigsaw puzzle (Noroozi and Favaro, 2016), and rotation prediction (Gidaris et al., 2018). However, these heuristic tasks might limit the generality of the learned representations (Chen et al., 2020). Recently, discriminative approaches based on contrastive learning have received increasing research attention and achieved prominent success in many computer vision tasks (Khosla et al., 2020). Contrastive learning operates by pulling similar samples (also called positive pairs) closer and pushing dissimilar samples (also called negative pairs) apart (Jaiswal et al., 2020). These methods have achieved results comparable to supervised learning methods on benchmark ImageNet dataset (Deng et al., 2009), such as the Simple framework for Contrastive Learning of visual Representations (SimCLR) (Chen et al., 2020), Swapping Assignments between multiple views of the same image (SwAV) (Caron et al., 2020), and Momentum Contrast (MoCo) (He et al., 2020). Regardless of the self-supervised approach used, semi-supervised learning (SSL) effectively builds upon model architectures pre-trained by a self-supervised approach method, that already learns knowledge from a large amount of unlabeled data. The SSL methods perform downstream tasks (e.g., images classification) involving a limited amount of labeled data (Lu et al., 2023). Recent studies demonstrate that these approaches outperform supervised learning methods in a variety of domain-specific computer vision tasks. For example, Muljo et al. (2023) utilized an SSL approach based on the SwAV algorithm to assist medical personnel in diagnosing COVID-19 infections, by classifying chest X-ray images into four categories: positive COVID-19, normal, lung opacity and viral pneumonia. Their SSL method demonstrated superior performance compared to supervised learning, achieving a 0.23 improvement in macro-averaged F1-score. Güldenring and Nalpantidis (2021) applied a similar SSL method to classify plants from agricultural images, obtaining a higher accuracy of up to 15%, compared to supervised learning approaches.

In this study, we proposed a two-stage semi-supervised learning method based on SimCLR to detect and quantify microparticles free from sludge and attached to sludge. We developed and evaluated the SSL method using microscopy images collected from the water lab of TU Delft. To the best of our knowledge, we are the first to propose and evaluate the SSL methods for identifying attachment between microparticles and sludge in biological wastewater treatment process.

2. The TUD-IPB dataset

We utilized the "TU Delft-Interaction between Particles and Biomass" (TUD-IPB) dataset, introduced in our previous study (Jia et al., 2024a). The data was collected from the experiments at the water lab of Delft University of Technology (TUD), Delft, the Netherlands. Fig. 1 shows the experimental procedure to build the TUD-IPB dataset. First, we collected the sludge mixture from a Nereda® sewage treatment plant (Utrecht, the Netherlands) with a treatment capacity of 74 700 m^3 d^{-1}. The sludge mixture was collected in a Nereda reactor after 30 min of aeration, then it was washed three times with mili-Q Water and sieved sequentially through meshes of 3.1 mm, 2.0 mm, 1.0 mm, and 0.2 mm (Fig. 1(a)). Second, we employed fluorescent microbeads to simulate sewage particles (Cospheric, WTW D-82362 Weilheim, Model: SEP 25, Order NO:209503, Diameter 250 μm, Density 1.02 g L^{-1}). These microbeads were ground into crushed particles (82.58 ± 47.95 μm) and diluted with mili-Q Water to prepare particle solution (2000–2500 particles L^{-1}) (Fig. 1 (b)). Third, we conducted the batch test with five 250 mL conical flasks. 200 mL particle solution was added in each flask, then the biomass fractions of >3.1 mm, 2.0–3.1 mm, 1.0–2.0 mm, 0.2–1.0 mm and <0.2 mm were added to each flask, respectively. The mixture was aerated for 60 min (Fig. 1(c)). Fourth, we collected 10 mL samples at 10 min, 30 min and 60 min from each flask, and observed these samples using a digital microscope (VHX-5000), and recorded microscopy images with a resolution of 1600 × 1200 pixel (Fig. 1(d)). Finally, we annotated microparticles free from biomass and entrapped in biomass in images with mask labels (Fig. 1(e)). Examples of images from this dataset are shown in Fig. A.1.

3. Methodology

3.1. Overview of the semi-supervised learning method

In this work, we proposed a two-stage semi-supervised learning method based on SimCLR for detecting and segmenting free particles and entrapped particles in microscopy images. It includes two stages: (1) self-supervised pre-training stage, and (2) supervised fine-tuning stage. Fig. 2 shows the outline of the SSL approach. In the first stage, we employed SimCLR to pre-train a ResNet50 network (He et al., 2016) on much unlabeled data. In the second stage, we constructed a Mask Region-based Convolutional Neural Network (Mask R-CNN) (He et al., 2017) for instance segmentation by adding additional deep learning layers after the pre-trained ResNet50. Then, we fine-tuned the Mask R-CNN on a small quantity of labeled data, to conduct a specific particle detection and segmentation downstream task in a supervised learning mode. We present the details of SimCLR and Mask R-CNN architecture in Sections 3.2 and 3.3, respectively. The details of the self-supervised pre-training stage and supervised fine-tuning stage are illustrated in Sections 3.4 and 3.5, respectively.

3.2. Simple framework for Contrastive Learning of visual Representations (SimCLR)

SimCLR is a simple self-supervised contrastive learning framework, that learns feature representations by pulling two augmented views of the same input image closer on the embedding space (Chen et al., 2020). Fig. 3 shows the schematic illustration of SimCLR. First, SimCLR employs data augmentation techniques to create two views (x_i and x_j) from the input image x. These two views are considered as a positive pair. In Fig. 3, we only show the cropping and resizing augmentation

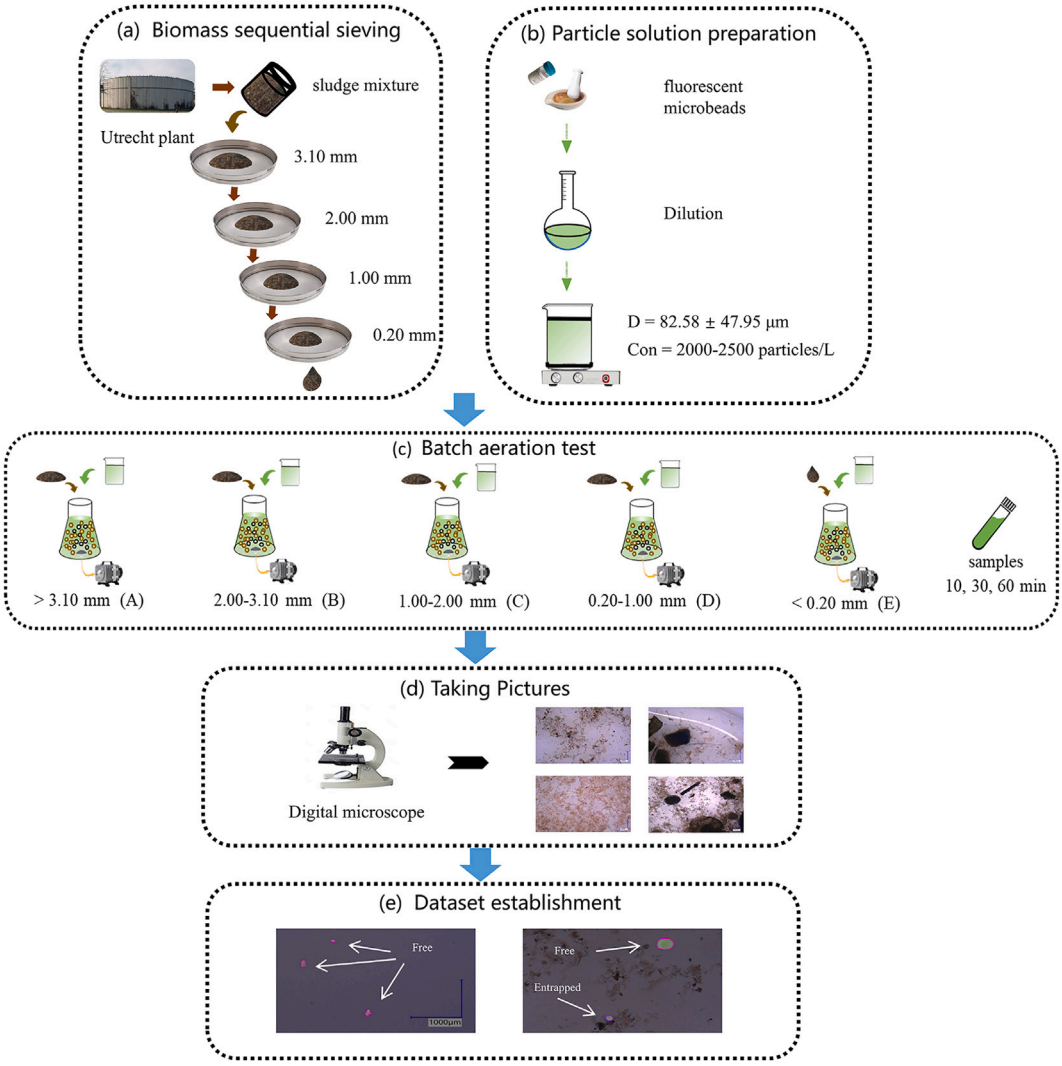


Fig. 1. Experimental procedure to build the TUD-IPB dataset.

technique. Second, a visual encoder network $f(\cdot)$ (e.g., a ResNet50 architecture) extracts feature representations (h_i and h_j) from these two augmented views. Third, SimCLR uses a projection head $g(\cdot)$ (e.g., 2-layer multilayer perceptron) to obtain z_i and z_j by mapping the feature representations (h_i and h_j) to the space. Finally, SimCLR learns data representations by maximizing agreement between two augmented views in the latent space via a contrastive loss. The contrastive loss function for this positive pair of examples (i, j) is given as follows:

$$\ell_{i,j} = -\log \frac{\exp\left(\mathrm{sim}\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k\neq i]} \exp\left(\mathrm{sim}\left(\boldsymbol{z}_i, \boldsymbol{z}_k\right)/\tau\right)} \tag{1}$$

$$sim(z_i, z_j) = \frac{z_i^{\top} z_j}{\|z_i\| \|z_j\|} \tag{2}$$

where $\mathbb{1}_{[k\neq i]} \in \{0, 1\}$ represents an indicator function equal to 1 if $k \neq i$; N is the number of input images in a mini-batch; τ is the temperature parameter; $sim(z_i, z_j)$ is the cosine similarity between z_i and z_j.

The careful selection of data augmentation techniques in SimCLR significantly affect model performance (Chen et al., 2020). The data augmentation techniques normally include geometric transformations and color space augmentations (Jia et al., 2023b; Shorten and Khoshgoftaar, 2019). In this study, we employed four commonly used data augmentation techniques in SimCLR: (1) random cropping and resizing; (2) horizontal flipping, (3) color distortion, and (4) Gaussian blur (Chen et al., 2020). Fig. 4 shows examples of each data augmentation technique. Cropping and resizing data augmentation technique involves randomly cropping an input image into two patches with a smaller resolution than the original input image and then resizing them into the same size (e.g., 224 × 224 pixel). Horizontal flipping involves reversing pixels of the image in the horizontal direction. The color distortion operation adjusts the brightness, contrast, and saturation of images. The Gaussian blur operation adjusts the blur strength.

3.3. Mask R-CNN for particle detection and segmentation

Fig. 5(a) illustrates the outline of the Mask R-CNN architecture with a ResNet backbone. It is a two-stage network, that extends the Faster R-CNN (Ren et al., 2015) for image segmentation by inserting a mask segmentation branch in parallel with the existing branch for predicting the bounding box of objects. It mainly includes two modules: (1) feature extraction, and (2) mask prediction and bounding-box prediction (i.e., classification and location prediction). In the first stage of the Mask R-CNN, the ResNet50 backbone (Fig. 5(b)) extracts feature maps from the input images. Then, the region proposal network (i.e., a fully convolutional network) creates a serious of region proposals from the shared feature maps. These proposals are considered as the possible Region of Interest (RoI). Together with the feature maps, these proposals are fed into the RoIAlign layer, that properly aligns the feature

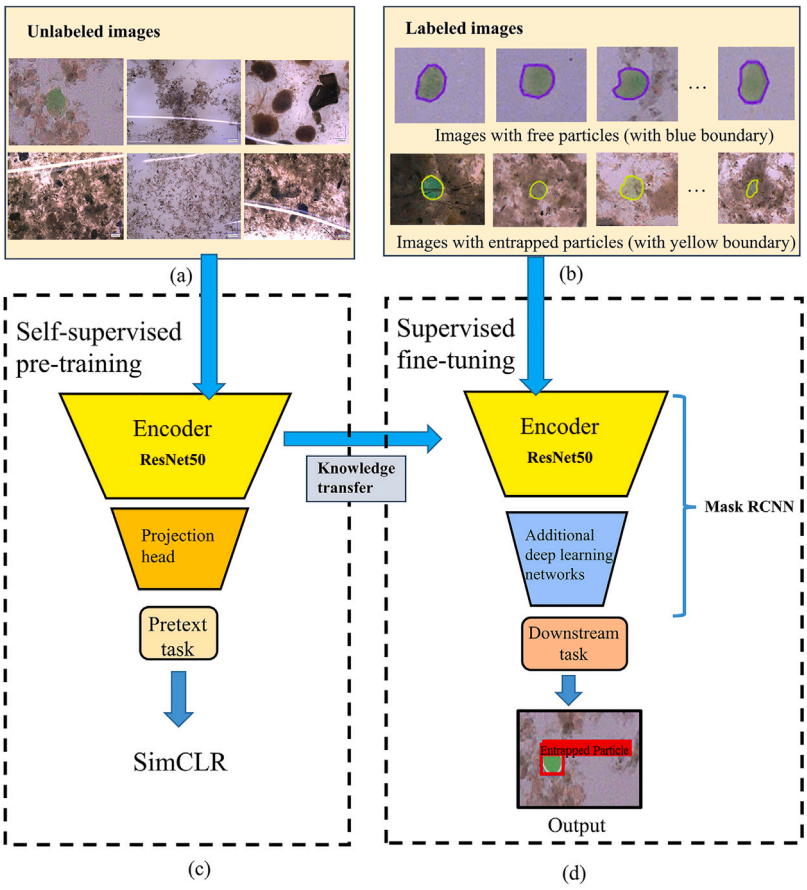


Fig. 2. The outline of the proposed two-stage semi-supervised learning method. In the first stage (c), we employed SimCLR to pre-train a ResNet50 encoder network with a projection head, using much unlabeled data (a). In the second stage (d), we created a Mask R-CNN architecture by adding extra deep learning layers to the ResNet50 backbone. Then, we fine-tuned this Mask R-CNN on a small quantity of labeled data (b), to perform a specific particle detection and segmentation downstream task. Note that the images shown in (b) are zoomed-in and cropped to improve visibility of particles, whereas the images used for model training and testing remain unaltered.

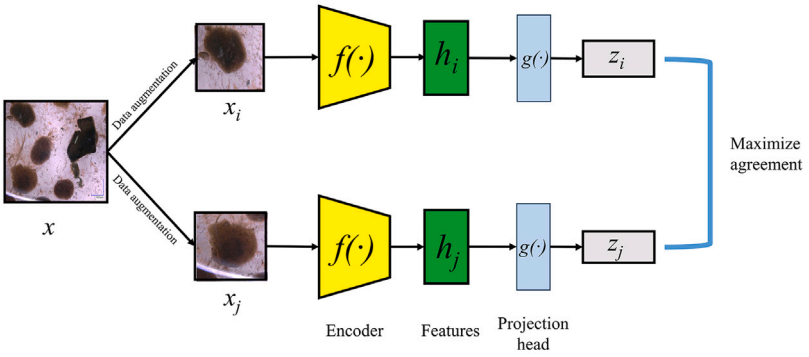


Fig. 3. The schematic illustration of SimCLR. First, SimCLR creates two correlated views (x_i and x_j) from the input image x using data augmentation techniques. Second, these views are processed by the encoder $f(\cdot)$ to extract feature representations (h_i and h_j). Third, SimCLR uses a projection head $g(\cdot)$ to obtain z_i and z_j by mapping the representations (h_i and h_j) to the latent space. Finally, SimCLR learns representations by maximizing agreement between two augmented views.

maps from each RoI with the input image and maps them into fixed-size feature maps. In the second stage, the mask head (i.e., convolutional layers) and detection head (i.e., fully connected layers) simultaneously predict the pixel-wise mask, and the bounding box with the class of the detected object from these standard-sized feature maps, respectively.

In this study, we used the ResNet50 as the backbone of the Mask R-CNN since it is the usually used model architecture in contrastive learning studies, due to its balance between size and learning capability (Jaiswal et al., 2020). The ResNet50 mainly includes two parts: (1) convolutional blocks Conv1 to Conv4, and (2) Conv5 (He et al., 2016). In the first stage of the SSL, both parts are pre-trained by SimCLR. In the second stage, we constructed the Mask R-CNN by employing Conv1 to Conv4 as the backbone and adding Conv5 after the RoIAlign layer.

3.4. Self-supervised pre-training with SimCLR

In the self-supervised pre-training stage (Fig. 2(c)), we initialized the ResNet50 backbone with pre-trained weights from ImageNet, subsequently fine-tuned the entire network using SimCLR on unlabeled data. This is also known as transfer learning, that provides a better initialization to the parameters of the network and prevents overfitting during model training (Wu et al., 2024). Initializing deep learning

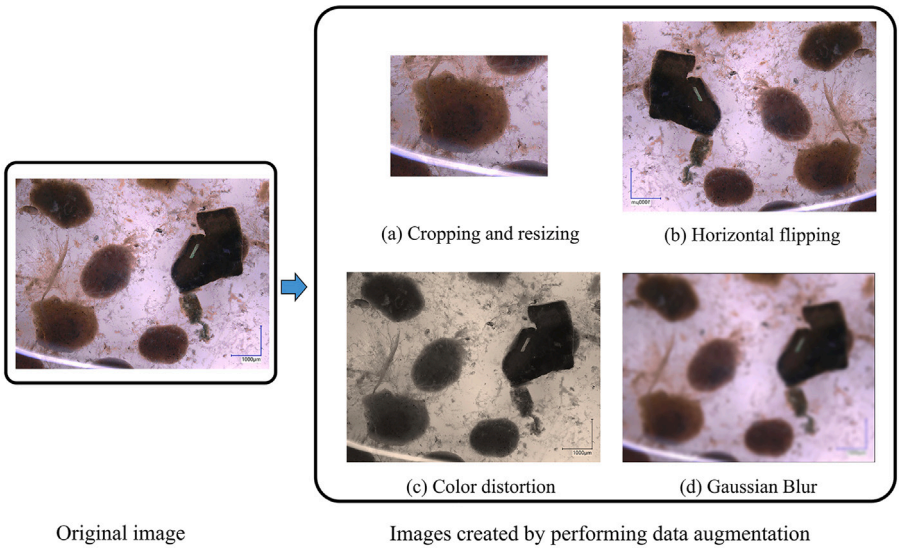


Fig. 4. Examples of data augmentation techniques in SimCLR. Left: an original microscopy image; Right: images generated by performing (a) cropping and resizing, (b) horizontal flipping, (c) color distortion, and (d) Gaussian blur. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

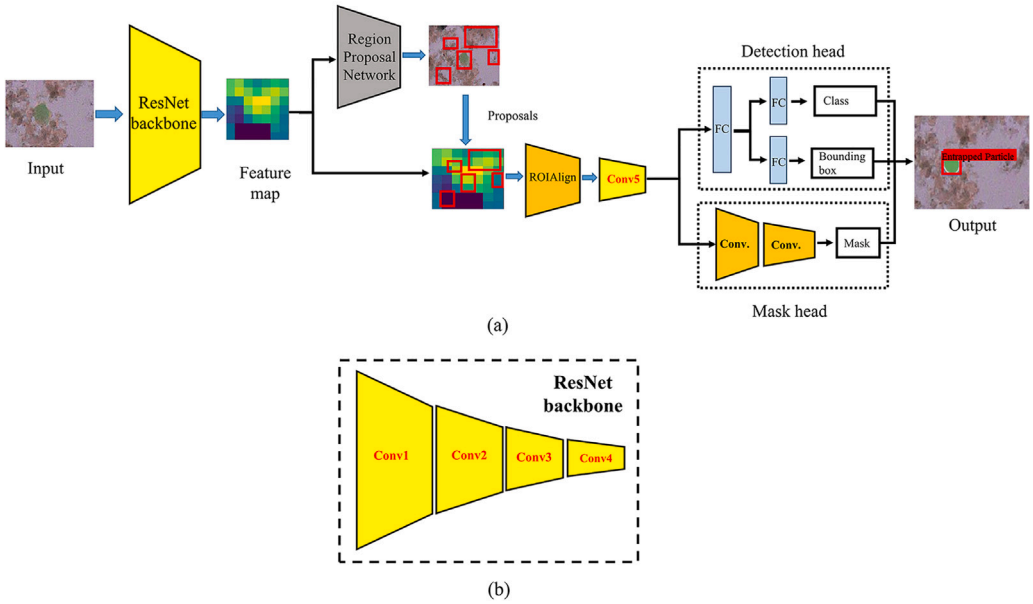


Fig. 5. The framework of the Mask R-CNN (a) with the ResNet backbone (b). The ResNet (yellow blocks) mainly include two modules: (1) convolutional blocks Conv1 to Conv4, and (2) Conv5. In the first stage of Mask R-CNN, the ResNet backbone extracts feature maps from the input images. From these feature maps, the region proposal network extracts region proposals as the possible Region of Interest (RoI). Then, these RoI and feature maps are fed into the RoIAlign layer to create standard-sized feature maps of each RoI. Finally, the mask head and detection head predict pixel-wise masks and bounding box from these feature maps, respectively. Acronyms used: Convolutional layer (Conv.), Fully connected layer (FC). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

models with ImageNet weights is a commonly employed method in computer vision tasks (Jia et al., 2024b). ImageNet weights employed in this work were obtained by training the ResNet50 on a subset (1.2 million images with 1000 categories) of the full ImageNet dataset.

3.5. Supervised fine-tuning: particle detection and segmentation

In the supervised fine-tuning stage (Fig. 2(d)), we first constructed a Mask R-CNN architecture by adding additional deep learning layers after the ResNet50 backbone obtained in the first stage. Then, we froze the first two convolutional blocks of the backbone (i.e., Conv1 and Conv2), and fine-tuned the Mask R-CNN on labeled data, to conduct the particle detection and segmentation downstream task. During the fine-tuning process, only the parameters of the unfrozen layers are updated, while those of the frozen layers are kept unchanged.

4. Experiment

To comprehensively evaluate the effectiveness and practical applicability of the SSL method, we conducted two experiments. Fig. 6 shows the flowchart of two experiments. In Experiment 1, we explored the potential of the SSL method for detecting and segmenting particles in microscopy images. For benchmarking, we compared the performance of this method against a baseline supervised learning approach. Furthermore, we analyzed how model performance varies with respect to the availability of labeled data for fine-tuning. Such analysis is vital for assessing the practical applicability of the SSL method in scenarios where annotated data is limited. In Experiment 2, we evaluated the practical applicability of the SSL method in monitoring the attachment between microparticles and sludge. With the detection results from the

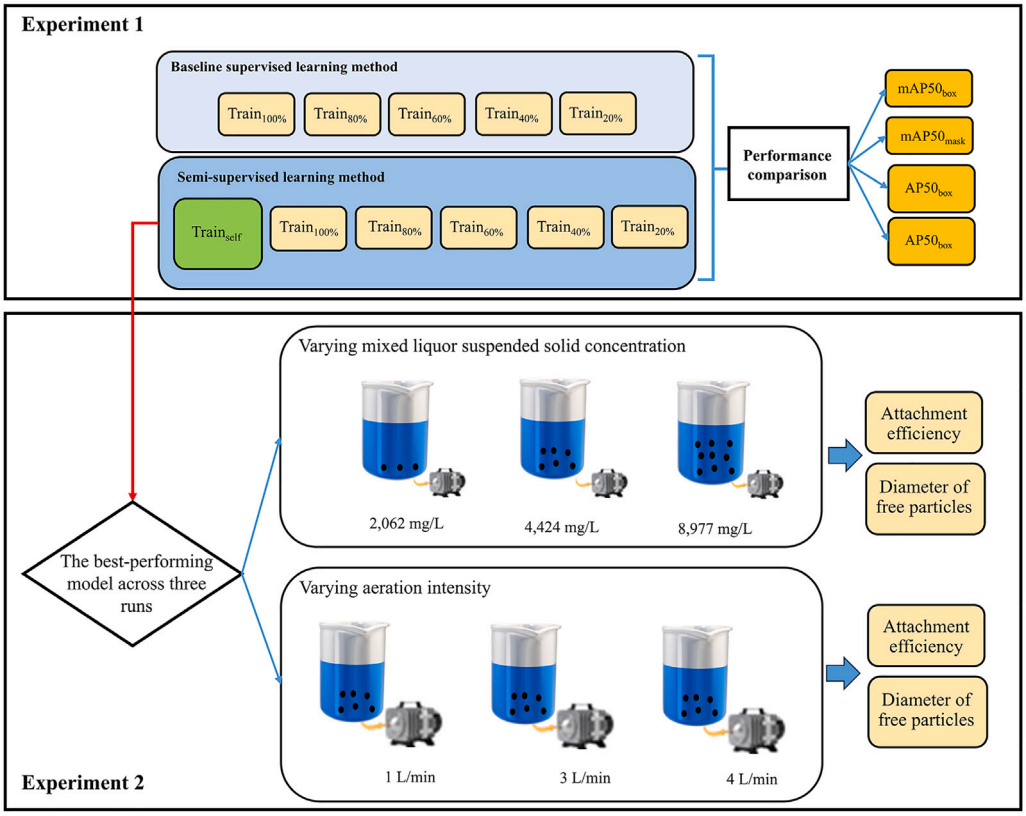


Fig. 6. The flowcharts of two experiments.

SSL methods, we investigated the influence of mixed liquor suspended solids (MLSS) concentration and aeration intensity on this attachment.

4.1. Experiment 1: SSL versus baseline methods

4.1.1. Dataset construction

Table 1 shows the subsets used in the first experiment to evaluate the effectiveness of the SSL and baseline method. First, we randomly selected 1284 images from the TUD-IPB dataset. Second, we created the $Train_{self}$ subset (1000 images) by randomly selecting approximately 80% of these images. These images are not annotated and are used for self-supervised pre-training. Third, we randomly split the remaining 284 images from the TUD-IPB dataset into $Train_{100\%}$, $Validation_{100\%}$ and Test subset in a ratio of 80:10:10. In these three subsets, we used mask labels to annotate the free particles and entrapped micro crushed particles for indicating their location and shape (see Fig. 2(b)). The $Train_{100\%}$, $Validation_{100\%}$ and Test subset includes 635, 51 and 63 particle items, respectively. The $Train_{100\%}$ subset is used for model fine-tuning. The $Validation_{100\%}$ subset is used to assess model performance during training and to prevent overfitting. The Test subset is used to evaluate the generalization capability of models on "unseen" data (Jia et al., 2023a).

To further evaluate the performance of the SSL method with regard to the availability of labeled data, we generated four smaller fine-tuning subsets by reducing the number of images of the $Train_{100\%}$ subset down to 20% at 20% intervals ($Train_{80\%}$ to $Train_{20\%}$), respectively. Similarly, we further generated four smaller validation subsets ($Validation_{80\%}$ to $Validation_{20\%}$).

4.1.2. Model development

We used subsets in Table 1 to develop SSL models and baseline supervised learning models. For developing SSL models, we first used SimCLR method to pre-train the ResNet50 encoder with a projection head on the $Train_{self}$ subset. Then, we fine-tuned the Mask R-CNN constructed based on the pre-trained ResNet50 backbone on all five fine-tuning subsets in a supervised learning manner (i.e., $Train_{100\%}$ to $Train_{20\%}$). To prevent overfitting, we validated the SSL models on the corresponding validation subset (i.e., $Validation_{100\%}$ to $Validation_{20\%}$). Finally, we assessed the model performance on Test subset.

The development of the baseline model only includes supervised fine-tuning procedure (see Fig. 2(d)), but does not include self-supervised pre-training procedure (see Fig. 2(c)). For developing baseline models, we employed the same Mask R-CNN with the ResNet50 backbone. For consistency, we also initialized the ResNet50 backbone with pre-trained weights from ImageNet. During the fine-tuning process, the first two blocks of the backbone (i.e., Conv1 and Conv2) are frozen, and the Mask R-CNN is subsequently fine-tuned on labeled data in a supervised manner. We employed the same datasets as the SSL models for fine-tuning, validation and testing the baseline models.

4.1.3. Performance assessment

To measure the performance of the Mask R-CNN for particle detection and segmentation, we used two common metrics: box-level mAP50 ($mAP50_{box}$) and mask-level mAP50 ($mAP50_{mask}$) (Jia et al., 2023a). The mAP50 denotes the mean average precision (mAP) of all categories with an intersection over union (IoU) threshold of 50%, which is defined as follows:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{3}$$

where AP_i is the average precision (AP) of the ith category for a given IoU threshold; n is the number of categories.

The IoU is the proportion of the intersection area of prediction and ground-truth to their union area, which is expressed as follows:

$$IoU = \frac{Area(Pred \cap GT)}{Area(Pred \cup GT)} \tag{4}$$

Table 1
Subsets in experiments.

Experiment	Dataset	Subset name	No. annotated entrapped particles	No. annotated free particles	No. total annotated particles	No. images
1	Training	Train-self	0	0	0	1000
		$Train_{100\%}$	209	426	635	226
		$Train_{80\%}$	165	353	518	180
		$Train_{60\%}$	127	246	373	135
		$Train_{40\%}$	82	180	262	91
		$Train_{20\%}$	44	73	117	46
	Validation	$Validation_{100\%}$	19	32	51	29
		$Validation_{80\%}$	13	30	43	23
		$Validation_{60\%}$	11	18	29	17
		$Validation_{40\%}$	8	12	20	12
		$Validation_{20\%}$	5	11	16	6
	Test	Test	24	39	63	29
2	$Test_{Exp2}$		0	0	0	322

where $Pred$ and GT are the predicted bounding box or mask and ground-truth bounding box or mask, respectively; A high IOU value indicates that the detected bounding box or mask is well-aligned with the ground-truth bounding box or mask.

The AP of one category is calculated as the area under the precision–recall curve, that is expressed as follows (Padilla et al., 2020):

$$AP = \int_0^1 p(r)dr \tag{5}$$

where $p(r)$ is the precision (p) at the recall of r.

The precision p and recall r are calculated using the number of true positives (TP), false positives (FP) and false negatives (FN), which are expressed as follows:

$$p = \frac{TP}{TP + FP} \tag{6}$$

$$r = \frac{TP}{TP + FN} \tag{7}$$

TP is the number of objects that are correctly classified and have an IoU equal or above the IoU threshold. FP represents the number of objects that are either misclassified or have an IoU below the IoU threshold. FN donates the number of ground-truth objects that are not detected by the model.

4.1.4. Implementation setup

For SimCLR pre-training, we pre-trained a ResNet50 with a projection head of 2-layer multilayer perceptron. We pre-trained it for 100 epochs with a batch size of 16 and a temperature of 0.1. We used an SGD optimizer with cosine annealing learning rate scheduling (Loshchilov and Hutter, 2016), with the initial rate of 0.01875 and the minimum value of 0. When fine-tuning SSL models and supervised learning models, we used a batch size of 4 and an SGD optimizer with a weight decay of 0.0001, a momentum of 0.9, and a fixed learning rate of 0.01. We fine-tuned the Mask R-CNN for 100 epochs and only saved the weights achieving the highest validation accuracy (i.e., $mAP50_{box}$) in validation subset. During testing, we adopt Non-Maximum Suppression algorithm with an IoU threshold of 0.5 to eliminate redundant (overlapping) bounding box and masks (Hosang et al., 2017). All models were implemented using Python programming language (version 3.8.16), Pytorch framework (version 1.8.1), as well as VISSL (Goyal et al., 2021) and Detectron2 (Wu et al., 2019) library. We developed models on a NVIDIA Tesla V100S GPU (32 GB) (Delft High Performance Computing Centre (DHPC), 2022).

To minimize the impact of randomization, we conducted the fine-tuning procedure three times for both SSL and supervised learning models, and only reported the mean values calculated from these three runs in Section 5. However, we conducted the analysis on the best-performing models out of the three runs, when discussing (1) the results of visual inspection on images predicted by SSL or supervised learning models in Experiment 1 (see Section 5.1), and (2) the post-processing results from DL models in Experiment 2 (see Section 5.2).

4.2. Experiment 2: Monitoring the attachment between microparticles and sludge

Various factors, such as MLSS concentration and aeration intensity, influence the performance of biological wastewater treatment (Yang et al., 2024). In this experiment, we applied the DL methods to detect this interaction from microscopy images. Then, we investigated the influence of these two factors on the attachment between microparticles and sludge based on the model's detection results.

We followed the preparation procedure of simulated wastewater outlined in Section 2. The microparticle concentration and diameter were 2000–2500 particles·L^{-1} and 82.58 ±47.95 μm, respectively. The biomass mixture was collected from the Utrecht wastewater treatment plant, and the fraction of 2.0–2.4 mm (pore size of the sieves) were used for this experiment. The experiment includes two groups: (1) MLSS concentration, and (2) aeration intensity. In each group, we conducted batch tests using three 500 mL conical flasks. In the first group, 20 mL, 40 mL, and 80 mL of granules were added to three separate flasks, respectively, followed by the addition of 200 mL simulated wastewater. The initial MLSS concentration of three flasks was 2062 mg L^{-1}, 4424 mg L^{-1} and 8977 mg L^{-1}, respectively. Each flask was then aerated for 60 min at an intensity of 2 L min^{-1}. In the second group, we added 40 mL granules into each flask, followed by the addition of 200 mL simulated wastewater. Three flasks were aerated for 60 min at different intensities: 1 L min^{-1} (uneven mixing), 3 L min^{-1} (intense mixing), and 4 L min^{-1} (super intense mixing), respectively. For both groups, we collected 10 mL samples from each flask at 10, 30, and 60 min. Each sample was filtrated by a 0.20 mm sieve, and the biomass solid on the sieve was rinsed slightly by miliQ Water. Then, we observed all the filtrate using a digital microscope (VHX-5000), and captured a total of 322 microscopy images to generate the $Test_{Exp2}$ dataset, as shown in Table 1. It is noted that these 322 images do not overlap with those used in Experiment 1 for model development. Finally, we used the best model developed in Experiment 1 to detect and segment particles in 322 images. We post-processed the results of model (i.e., the number and diameter of detected particles) to measure (1) the attachment efficiency of microparticles to sludge, and (2) the average diameter of free particles under different experiment conditions. Attachment efficiency is calculated as the percentage of the number of the attached microparticles respect to the total number of microparticles in raw water. Particle diameter is defined as the diameter of a circle with an area equal to that of the particle. Lengths were calculated using a scale bar to convert pixels to μm.

5. Results and discussion

5.1. Experiment 1: SSL versus baseline methods

Fig. 7 compares the mean accuracy in $mAP50_{box}$ and $mAP50_{mask}$ of the SSL and baseline methods on the Test subset. All results calculated

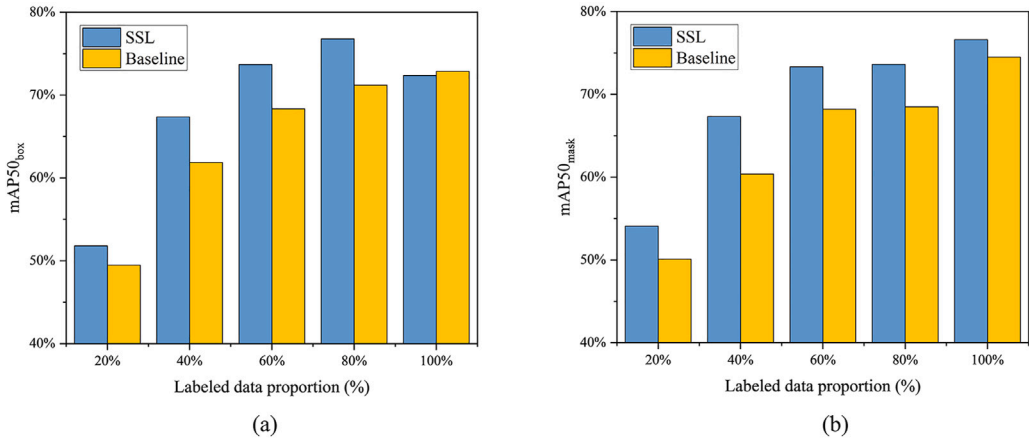


Fig. 7. Model performance of the SSL and baseline methods on the Test subset with different proportion of labeled data for fine-tuning. Performance metrics include (a) $mAP50_{box}$ and (b) $mAP50_{mask}$.

from three runs are shown in Tables B.1 and B.2. The results show that the SSL method outperforms the baseline method in both $mAP50_{box}$ and $mAP50_{mask}$ in most cases. The SSL method achieves an $mAP50_{box}$ varying between 51.8% to 76.8% and an $mAP50_{mask}$ ranging from 54.1% to 76.6%. The baseline method obtains an $mAP50_{box}$ varying between 49.5% to 72.9% and an $mAP50_{mask}$ ranging from 50.1% to 74.5%. Moreover, we observed that the proposed SSL approach significantly outperforms the baseline method with an improvement in both $mAP50_{box}$ and $mAP50_{mask}$ of over 5%, when few labeled images are available for training in a supervised manner (i.e., $Train_{40\%}$ to $Train_{80\%}$). For example, when models are fine-tuned on 40% of labeled images, the SSL method improves the $mAP50_{box}$ of 5.5% and the $mAP50_{mask}$ of 6.9%, compared with the baseline method. The SSL method achieves a slight improvement in $mAP50_{box}$ of 2.3% and $mAP50_{mask}$ of 4.0% compared with the baseline method, even if models are fine-tuned on insufficient labeled images ($Train_{20\%}$). When a relatively larger number of labeled images are available for fine-tuning (i.e., $Train_{100\%}$), these two methods performs similarly in $mAP50_{box}$, while the SSL method slightly improves the $mAP50_{mask}$ of 2.1%, compared to the baseline method.

We found that the SSL method could achieve competitive performance with the availability of less labeled images, compared with the baseline method. For instance, the SSL method with 40% of labeled images for fine-tuning (91 images with 262 annotated particles) obtains an $mAP50_{mask}$ of 67.3%, similar to that ($mAP50_{mask}$ = 68.5%) achieved by the baseline method with 80% of labeled images for fine-tuning (180 images with 518 annotated particles). These results indicate that the features representations learned by SimCLR on context-related images is more effective than ImageNet representations for particle detection and segmentation downstream task, while the ImageNet dataset used in this study includes over 1000 times more images than the $Train_{self}$ subset with context-related images for self-supervised pre-training.

Fig. 7 also illustrates that both the SSL and supervised learning method yield a sharp increase in both $mAP50_{box}$ and $mAP50_{mask}$ as the number of labeled images for fine-tuning increases from 46 ($Train_{20\%}$) to 91 ($Train_{40\%}$). For example, the SSL method yields an improvement in $mAP50_{box}$ of 15.6%, when expanding fine-tuning dataset size from 46 images to 91 images. Compared to other fine-tuning subsets, the $Train_{20\%}$ subset do not include a sufficient amount of labeled data, resulting in $mAP50_{box}$ and $mAP50_{mask}$ below 60% achieved by models.

Table 2 compares the detection and segmentation performance of the SSL and baseline methods for each class. The results show that the SSL significantly outperforms baseline method on entrapped particle detection and segmentation in most cases. The SSL method achieves an $AP50_{box}$ ranging from 29.5% to 65.3% and an $AP50_{mask}$ ranging from 30.3% to 63.9%. The baseline method obtains an $AP50_{box}$ varying between 15.0% and 57.5% and an $AP50_{mask}$ varying between 15.1% and 60.7%. When few labeled images are available for fine-tuning (i.e., $Train_{20\%}$ to $Train_{80\%}$), the SSL method yields an improvement in $AP50_{box}$ ranging from 12.1% to 16.5% and $AP50_{mask}$ ranging from 12.7% to 19.2%, compared to the baseline method. For free particle detection and segmentation, the SSL method performs worse than the baseline method in most cases. The SSL method obtains an $AP50_{box}$ ranging from 74.1% to 89.3% and an $AP50_{mask}$ ranging from 77.9% to 89.3%, while the baseline method achieves an $AP50_{box}$ ranging from 83.9% to 91.0% and an $AP50_{mask}$ ranging from 85.1% to 91.0%. When little labeled data is available for model fine-tuning (i.e., $Train_{20\%}$ to $Train_{80\%}$), the baseline method yields an improvement in $AP50_{box}$ ranging from 1.4% to 9.8% and $AP50_{mask}$ ranging from 2.5% to 7.3%, compared to the SSL method.

Fig. 8 shows some examples of bounding boxes and masks predicted by the Mask R-CNN on the Test subset using the SSL and baseline method. More examples can be found in Fig. B.1. This figure shows the prediction results of the models fine-tuned on the $Train_{80\%}$ subset. As shown in this figure, both the SSL and baseline method can correctly detect some particles, such as free particles in Fig. 8(a) and (b), and entrapped particles in Fig. 8(c). However, compared the SSL method, the baseline method falsely identifies a free particle as an entrapped particle in Fig. 8(b). Both methods yield few FPs. For example, the SSL method falsely detects background pixels as free particles in Fig. 8(b), and the baseline method falsely detects background pixels as free particles in Fig. 8(a) and entrapped particles in Fig. 8(c).

5.2. Experiment 2: Monitoring the attachment between microparticles and sludge

Fig. 9 shows the attachment efficiency of microparticles to sludge, and the average diameter of free particles under varying MLSS concentration and aeration intensity. These results are measured by post-processing the detection results of Mask R-CNN using the SSL method. This Mask R-CNN model is fine-tuned on the $Train_{80\%}$ subset in Experiment 1. We found that the average attachment efficiency of microparticles decreases with the increase of MLSS concentration (see Fig. 9(a)). When the initial MLSS concentration is 2062 mg L^{-1}, 4424 mg L^{-1} and 8977 mg L^{-1}, the average attachment efficiencies are 66.28 ± 3.11% and 51.18 ± 22.48% and 39.12 ± 21.21%, respectively. This indicates that the adsorption between microparticles and sludge is not determined by the surface area. Although increasing MLSS concentration provides more surface area, it also enhances the collision between sludge, resulting in detach of already attached microparticles (Tijhuis et al., 1994). Since bigger micro particles are difficult to be attached by sludge, better attachment efficiency is associated with

Table 2
The performance of the SSL and baseline method for each class.

Fine-tuning subset	Method	Test accuracy			
		$AP50_{box}$ per class		$AP50_{mask}$ per class	
		Entrapped particle	Free particle	Entrapped particle	Free particle
$Train_{20\%}$	SSL	**29.5%**	74.1%	**30.3%**	77.9%
	Baseline	15.0%	**83.9%**	15.1%	**85.1%**
$Train_{40\%}$	SSL	**50.7%**	84.2%	**50.7%**	84.0%
	Baseline	34.2%	**89.3%**	31.5%	**89.3%**
$Train_{60\%}$	SSL	**60.6%**	86.7%	**59.9%**	86.7%
	Baseline	48.6%	**88.1%**	47.2%	**89.3%**
$Train_{80\%}$	SSL	**65.3%**	88.3%	**58.9%**	88.3%
	Baseline	51.3%	**91.0%**	46.0%	**91.0%**
$Train_{100\%}$	SSL	55.4%	**89.3%**	**63.9%**	**89.3%**
	Baseline	**57.5%**	88.3%	60.7%	88.3%

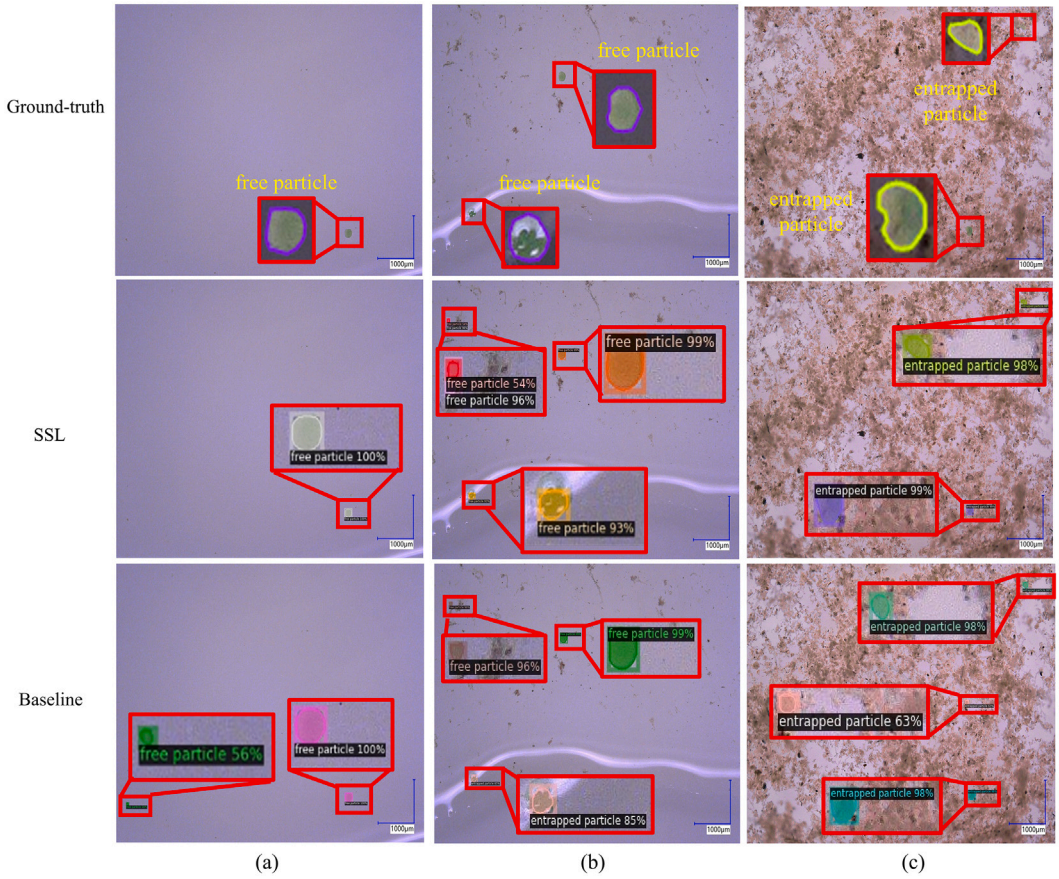


Fig. 8. Examples of bounding boxes and masks predicted by the Mask R-CNN on the Test subset using SSL and baseline method. The models were fine-tuned on the $Train_{80\%}$ subset.

lower diameter of free particles (Ranzinger et al., 2020). The intensified collision results in more bigger free microparticles in the mixture. When the initial MLSS concentration is 2062 mg L^{-1}, 4424 mg L^{-1} and 8977 mg L^{-1}, the average diameter of free particles during aeration is 76.08 ± 31.49 μm, 87.64 ± 39.41 μm and 101.69 ± 38.93 μm, respectively. In contrast, increasing aeration intensity has a limited effect on the attachment efficiency between microparticles and sludge (see Fig. 9(b)). When the initial MLSS concentration is similar (2924–3102 mg L^{-1}), the average attachment efficiency of microparticles under aeration intensities of 1 L min^{-1}, 3 L min^{-1}, and 4 L min^{-1} is 17.74 ± 9.93%, 23.90 ± 13.95% and 16.14 ± 14.89%, respectively. Although the hydraulic shear disturbance caused by different aeration intensity is different (Remmas et al., 2017), its effect on attachment efficiency is limited. Since the net attachment is determined by the balance between attachment and detachment. In this experiment, the aeration lasted for 60 min, the attachment efficiency under low aeration intensity (1 L min^{-1}) continuously increases, while the attachment efficiency under high aeration intensity (4 L min^{-1}) continuously decreases. Although higher aeration intensity indeed increases attachment by enhancing mass transfer initially, it also increases the collision probability to detach the microparticles (Bruckner et al., 2011). The average diameter of free microparticles is also similar (77.78 ± 19.38 μm,86.91 ± 27.90 μm and 87.40 ± 42.09 μm, respectively), which verifies that the aeration intensity has a limited effect on the attachment efficiency.

In the actual biological wastewater treatment process, to ensure that more influent microparticles are utilized by microorganisms, enhancing attachment by appropriately reducing the MLSS concentration has

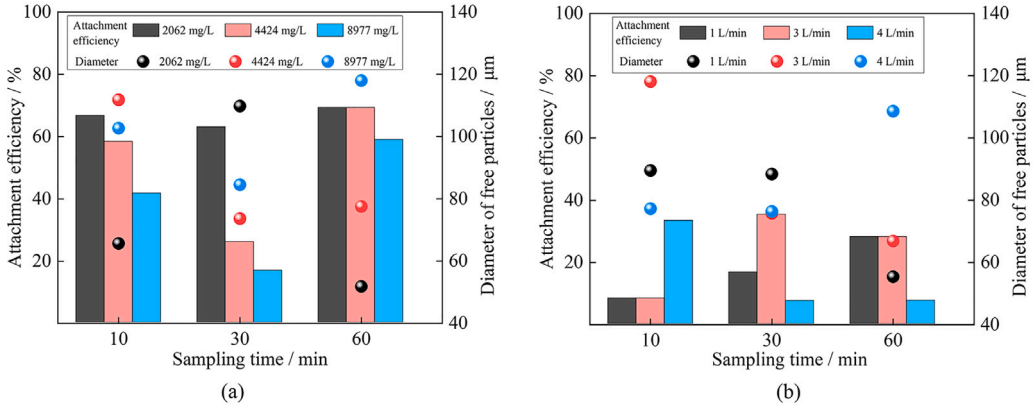


Fig. 9. The attachment efficiency of microparticles to sludge and the average diameter of free particles under varying (a) mixed liquor suspended solids (MLSS) concentration and (b) aeration intensity.

proven to be effective. Furthermore, if the aeration duration is sufficiently long, using a lower aeration intensity is suggested to achieve both high attachment efficiency and energy savings.

5.3. Implications for biological wastewater treatment

The biomass mixture used in this study is consisted of both granular sludge and conventional activated sludge flocs, and it covers the size of microbial aggregates commonly used in wastewater treatment (De Kreuk et al., 2007). Given the minimal color variation among the microbial aggregates of actual municipal wastewater treatment plants (typically appearing as dark brown), and the widespread use of fluorescent microbeads as a tracer (Sorasan et al., 2022), our developed model has the significant potential to be applied in biological wastewater treatment system.

On the one hand, the model helps to assess the fate of microplastics in wastewater treatment plants. It can investigate the attachment and detachment between microplastics and sludge, thereby elucidating the transfer behavior of microplastics within biological wastewater treatment systems. The distribution of microplastics between effluent and residual sludge depends on some factors, including particle size, shape, and surface properties (Wang et al., 2023). If microplastics predominantly accumulate in the residual sludge, incineration would be a suitable sludge treatment method; otherwise, anaerobic digestion would be more appropriate (Ottosen et al., 2016). On the other hand, the model allows to evaluate the removal performance of particulate matter in an affordable manner for the long-term monitoring. Many substances in actual wastewater exist in particulate form. For biodegradable particulate matter, they need to undergo attachment, hydrolysis, and other steps before being utilized by microorganisms. In contrast, the nonbiodegradable particulate matter only exhibits attachment and detachment with sludge. The model can help to investigate the removal performance of particulate matter within biological wastewater treatment systems by microscopy images, thereby improving its attachment performance through optimizing operational parameters.

5.4. Limitation and future works

The SSL methods we proposed consistently outperform the conventional supervised learning approaches in the detection and segmentation of free and entrapped particles from microscopy images. Additionally, the results indicate the practical applicability of the SSL methods in identifying the attachment between microparticles and sludge. However, this study only focused on the interactions between particulate matter and the granular sludge within the size range of 2.0–2.4 mm, whereas the biomass size distribution in the Nereda® system is considerably broader (Pronk et al., 2015). Future research should explore the interactions between particulate matter and granules of varying sizes. Furthermore, the activated sludge process is the most widely used technology in wastewater treatment, typically involving sludge flocs smaller than 200 μm. The SSL methods could be applied to examine interactions between particulate matter and activated sludge, focusing on effects of aeration rate, sludge concentration, reaction time, and mechanical stirring intensity.

Meanwhile, we identify several limitations in our methodology and dataset, and outline future works needed to facilitate real-world applications. First, the dataset for self-supervised pre-training contains much fewer images compared to large-scale comprehensive datasets, such as ImageNet. Future work should focus on enhancing model performance by expanding the scale of the pre-training dataset. Literature demonstrates significant benefits for SSL from this implementation. For example, Goyal et al. (2022) showed that the performance of the SSL method continuously improves with an increase in the size of the pre-training dataset, ranging from 1.2 million to 14 million to 1 billion. Second, the experiments in this study lacks hyper-parameter optimization and a systematic evaluation of the benefits derived from the selection of data augmentation techniques in SimCLR, due to computational resource limitations. Lastly, we did not explore the effectiveness of other contrastive learning methods due to computational resource constraints, such as SwAV (Caron et al., 2020) and MoCo (He et al., 2020). More focus on these aspects are needed to obtain a robust model.

6. Conclusions

Monitoring the particle transport process in biological wastewater treatment system is essential for enhancing the treatment performance. To detect and quantify microparticles free from sludge and entrapped in sludge, supervised deep learning algorithms in the field of computer vision are promising techniques. However, supervised learning algorithms rely on a large number of carefully labeled samples for training. Moreover, the labeling work is labor-intensive and costly, and relies on domain-specific knowledge. This hinders obtaining notable model generalization capability, underpinning the development of robust computer vision systems for structural monitoring of particle transport process. To address this concern, we proposed a semi-supervised learning (SSL) method for detecting microparticle transfer process in biological wastewater treatment systems. First, we extracted feature representations from unlabeled images by a self-supervised learning method SimCLR. Then, we fine-tuned the model on a limited amount of labeled data in a supervised learning manner.

To evaluate the performance of the proposed SSL methods, we conducted an experiment using microscopy images collected from the water lab of TU Delft. We compared the performance of the SSL methods to the conventional supervised learning methods with ImageNet

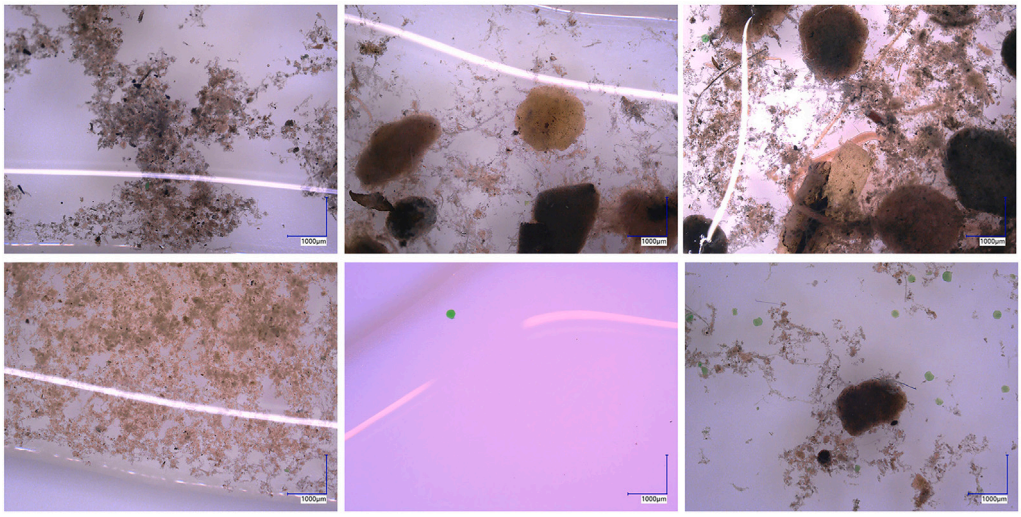


Fig. A.1. Examples of microscopy images collected from the experiments at the water lab of TU Delft.

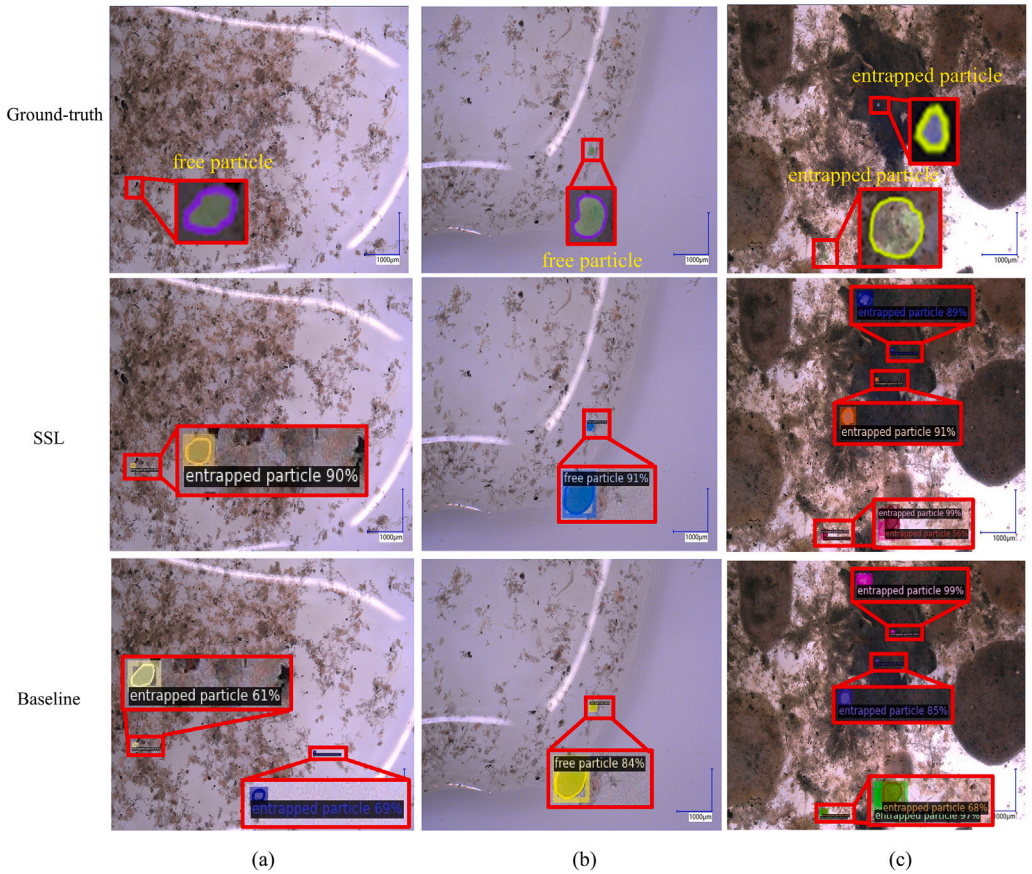


Fig. B.1. Detection and segmentation results of the SSL and baseline method on the Test subset. The models were fine-tuned on the $\text{Train}_{80\%}$ subset.

pre-trained weights. The results show that the SSL methods achieve a significant improvement in box-level and mask-level mean average precision (mAP_{box} and mAP_{mask}) of over 5% compared to the supervised learning approaches in scenarios with limited available labeled data. Especially, it yields a significant improvement in mAP_{box} and mAP_{mask} of over 12% with regard to entrapped particle detection and segmentation. Our findings indicate that the feature representations obtained through self-supervised learning on domain-specific images may prove superior to those derived from pre-training on much larger, but unrelated, datasets. Based on the detection results from the SSL methods, we measured the attachment efficiency of microparticles to sludge under varying mixed liquor suspended solids concentration and aeration intensity. The accurate measurement results indicate the practical applicability of the SSL methods in identifying the attachment between microparticles and sludge.

Our proposed SSL methods reduces dependency on the availability of labeled data, yet achieves competitive performance. These methods are especially suitable for cases where collecting data is relatively

Table B.1
The performance of the semi-supervised learning method across three runs.

Fine-tuning subset	Fine-tuning run	Fine-tuning time (min)	Validation accuracy ($mAP50_{box}$)	Test accuracy					
				$mAP50_{box}$	$mAP50_{mask}$	$AP50_{box}$ per class		$AP50_{mask}$ per class	
						Entrapped particle	Free particle	Entrapped particle	Free particle
$Train_{20\%}$	1	266	67.7%	49.8%	49.8%	30.2%	69.5%	30.2%	69.5%
	2	365	74.8%	51.0%	54.9%	32.2%	69.7%	28.8%	80.9%
	3	391	74.6%	54.6%	57.6%	26.0%	83.2%	32.0%	83.2%
	Average	341	72.4%	51.8%	54.1%	29.5%	74.1%	30.3%	77.9%
$Train_{40\%}$	1	304	88.8%	68.1%	68.1%	49.3%	87.0%	49.3%	87.0%
	2	431	88.8%	64.7%	66.9%	46.8%	82.7%	51.7%	82.2%
	3	445	87.6%	69.3%	67.0%	55.9%	82.8%	51.1%	82.8%
	Average	393	88.4%	67.4%	67.3%	50.7%	84.2%	50.7%	84.0%
$Train_{60\%}$	1	342	87.6%	76.2%	76.2%	63.4%	89.0%	63.4%	89.0%
	2	487	87.4%	74.5%	74.5%	58.7%	90.2%	58.7%	90.2%
	3	502	86.1%	70.4%	69.3%	59.8%	81.0%	57.5%	81.0%
	Average	444	87.0%	73.7%	73.3%	60.6%	86.7%	59.9%	86.7%
$Train_{80\%}$	1	373	89.2%	79.9%	73.7%	69.8%	89.9%	57.4%	89.9%
	2	540	89.7%	76.0%	72.4%	61.7%	90.4%	54.5%	90.4%
	3	522	90.2%	74.5%	74.8%	64.3%	84.7%	64.9%	84.7%
	Average	478	89.7%	76.8%	73.6%	65.3%	88.3%	58.9%	88.3%
$Train_{100\%}$	1	348	74.0%	77.1%	77.1%	67.0%	87.2%	67.0%	87.2%
	2	514	75.2%	71.6%	75.8%	54.0%	89.2%	62.4%	89.2%
	3	502	75.4%	68.4%	77.0%	45.3%	91.5%	62.4%	91.5%
	Average	455	74.9%	72.4%	76.6%	55.4%	89.3%	63.9%	89.3%

Table B.2
The performance of the baseline method across three runs.

Fine-tuning subset	Fine-tuning run	Fine-tuning time (min)	Validation accuracy ($mAP50_{box}$)	Test accuracy					
				$mAP50_{box}$	$mAP50_{mask}$	$AP50_{box}$ per class		$AP50_{mask}$ per class	
						Entrapped particle	Free particle	Entrapped particle	Free particle
$Train_{20\%}$	1	261	79.2%	50.3%	52.2%	17.0%	83.5%	17.2%	87.2%
	2	400	80.5%	50.1%	50.1%	14.1%	86.1%	14.1%	86.1%
	3	425	74.7%	48.0%	48.0%	13.9%	82.1%	13.9%	82.1%
	Average	362	78.1%	49.5%	50.1%	15.0%	83.9%	15.1%	85.1%
$Train_{40\%}$	1	300	89.1%	65.3%	60.4%	38.0%	92.6%	28.3%	92.6%
	2	453	88.3%	59.6%	59.6%	30.5%	88.7%	30.5%	88.7%
	3	440	87.8%	60.3%	61.2%	34.0%	86.6%	35.8%	86.6%
	Average	398	88.4%	61.7%	60.4%	34.2%	89.3%	31.5%	89.3%
$Train_{60\%}$	1	338	90.5%	72.1%	73.1%	55.8%	88.4%	55.8%	90.4%
	2	462	89.7%	62.7%	61.3%	38.5%	86.9%	34.3%	88.4%
	3	441	91.5%	70.2%	70.2%	51.4%	89.0%	51.4%	89.0%
	Average	414	90.6%	68.3%	68.2%	48.6%	88.1%	47.2%	89.3%
$Train_{80\%}$	1	369	91.6%	69.0%	66.2%	48.8%	89.2%	43.2%	89.2%
	2	464	91.2%	69.2%	66.5%	46.1%	92.3%	40.7%	92.3%
	3	450	91.1%	75.4%	72.8%	59.1%	91.6%	54.1%	91.6%
	Average	428	91.3%	71.2%	68.5%	51.3%	91.0%	46.0%	91.0%
$Train_{100\%}$	1	344	81.3%	77.7%	76.6%	63.5%	91.9%	61.3%	91.9%
	2	428	72.9%	72.4%	72.4%	54.6%	90.2%	54.6%	90.2%
	3	445	74.9%	68.5%	74.5%	54.3%	82.8%	66.3%	82.8%
	Average	406	76.4%	72.9%	74.5%	57.5%	88.3%	60.7%	88.3%

easy, while labeling data manually is costly and labor-intensive. We suggest stakeholders using these methods to develop more robust deep learning models for long-term monitoring of particle transfer process in biological wastewater treatment systems.

CRediT authorship contribution statement

Tianlong Jia: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jing Yu:** Software, Methodology. **Ao Sun:** Data curation. **Yipeng Wu:** Data curation. **Shuo Zhang:** Data curation. **Zhaoxu Peng:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Author Ao Sun is employed by the company Power China Zhongnan Engineering Corporation Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

The work is supported by China Scholarship Council, China (No. 202006160032; 202107045010). We acknowledge the use of computational resources from Delft High Performance Computing Centre (https://www.tudelft.nl/dhpc).

Appendix A. Dataset information

See Fig. A.1.

Appendix B. Experiment results

See Tables B.1 and B.2 and Fig. B.1.

Data availability

The code for this study is available on https://github.com/TianlongJia/deep_pollutant_SSL. The data used in this study including images and mask annotations are available for download from Zenodo at https://doi.org/10.5281/zenodo.13374998.

References

Alvi, M., Batstone, D., Mbamba, C.K., Keymer, P., French, T., Ward, A., Dwyer, J., Cardell-Oliver, R., 2023. Deep learning in wastewater treatment: a critical review. Water Res. 120518.

Alvim, C.B., Bes-Piá, M., Mendoza-Roca, J.A., 2020. Separation and identification of microplastics from primary and secondary effluents and activated sludge from wastewater treatment plants. Chem. Eng. J. 402, 126293.

Benson, N.U., Agboola, O.D., Fred-Ahmadu, O.H., De-la Torre, G.E., Oluwalana, A., Williams, A.B., 2022. Micro (nano) plastics prevalence, food web interactions, and toxicity assessment in aquatic organisms: a review. Front. Mar. Sci. 9, 851281.

Blackburn, K., Green, D., 2022. The potential effects of microplastics on human health: What is known and what is unknown. Ambio 51 (3), 518–530.

Bruckner, C.G., Rehm, C., Grossart, H.-P., Kroth, P.G., 2011. Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. Environ. Microbiol. 13 (4), 1052–1063.

Campbell, K., Wang, J., Daniels, M., 2019. Assessing activated sludge morphology and oxygen transfer performance using image analysis. Chemosphere 223, 694–703.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A., 2020. Unsupervised learning of visual features by contrasting cluster assignments. Adv. Neural Inf. Process. Syst. 33, 9912–9924.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Cydzik-Kwiatkowska, A., Zielińska, M., Bernat, K., Bułkowska, K., Wojnowska-Baryła, I., 2020. Insights into mechanisms of bisphenol a biodegradation in aerobic granular sludge. Bioresour. Technol. 315, 123806.

De Kreuk, M., Kishida, N., Van Loosdrecht, M., 2007. Aerobic granular sludge–state of the art. Water Sci. Technol. 55 (8–9), 75–81.

Delft High Performance Computing Centre (DHPC), 2022. DelftBlue supercomputer (phase 1). https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase1.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.

Doersch, C., Gupta, A., Efros, A.A., 2015. Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1422–1430.

Feng, Y., Zhang, Y., Quan, X., Chen, S., 2014. Enhanced anaerobic digestion of waste activated sludge digestion by the addition of zero valent iron. Water Res. 52, 242–250.

Gidaris, S., Singh, P., Komodakis, N., 2018. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Goyal, P., Duval, Q., Reizenstein, J., Leavitt, M., Xu, M., Lefaudeux, B., Singh, M., Reis, V., Caron, M., Bojanowski, P., Joulin, A., Misra, I., 2021. VISSL. https://github.com/facebookresearch/vissl.

Goyal, P., Duval, Q., Seessel, I., Caron, M., Misra, I., Sagun, L., Joulin, A., Bojanowski, P., 2022. Vision models are more robust and fair when pretrained on uncurated images without supervision. arXiv preprint arXiv:2202.08360.

Güldenring, R., Nalpantidis, L., 2021. Self-supervised contrastive learning on agricultural images. Comput. Electron. Agric. 191, 106510.

He, K., Fan, H., Wu, Y., Xie, S., Girshick, R., 2020. Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9729–9738.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hosang, J., Benenson, R., Schiele, B., 2017. Learning non-maximum suppression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4507–4515.

Hu, X., Chen, Y., 2024. Response mechanism of non-biodegradable polyethylene terephthalate microplastics and biodegradable polylactic acid microplastics to nitrogen removal in activated sludge system. Sci. Total Environ. 917, 170516.

Jaiswal, A., Babu, A.R., Zadeh, M.Z., Banerjee, D., Makedon, F., 2020. A survey on contrastive self-supervised learning. Technol. 9 (1), 2.

Jia, T., Kapelan, Z., de Vries, R., Vriend, P., Peereboom, E.C., Okkerman, I., Taormina, R., 2023a. Deep learning for detecting macroplastic litter in water bodies: a review. Water Res. 119632.

Jia, T., Peng, Z., Yu, J., Piaggio, A.L., Zhang, S., de Kreuk, M.K., 2024a. Detecting the interaction between microparticles and biomass in biological wastewater treatment process with deep learning method. Sci. Total Environ. 175813.

Jia, T., Vallendar, A.J., de Vries, R., Kapelan, Z., Taormina, R., 2023b. Advancing deep learning-based detection of floating litter using a novel open dataset. Front. Water 5, 1298465.

Jia, T., de Vries, R., Kapelan, Z., van Emmerik, T.H., Taormina, R., 2024b. Detecting floating litter in freshwater bodies with semi-supervised deep learning. Water Research 266, 122405.

Jing, L., Tian, Y., 2020. Self-supervised visual feature learning with deep neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (11), 4037–4058.

Keller, A.S., Jimenez-Martinez, J., Mitrano, D.M., 2019. Transport of nano-and microplastic through unsaturated porous media from sewage sludge application. Environ. Sci. Technol. 54 (2), 911–920.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D., 2020. Supervised contrastive learning. Adv. Neural Inf. Process. Syst. 33, 18661–18673.

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Laskar, N., Kumar, U., 2019. Plastics and microplastics: A threat to environment. Environ. Technol. Innov. 14, 100352.

Li, Y.-Q., Zhao, B.-H., Zhang, Y.-Q., Zhang, X.-Y., Chen, X.-T., Yang, H.-S., 2023. Effects of polyvinylchloride microplastics on the toxicity of nanoparticles and antibiotics to aerobic granular sludge: Nitrogen removal, microbial community and resistance genes. Environ. Res. 238, 117151.

Loshchilov, I., Hutter, F., 2016. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983.

Lu, M.Y., Chen, B., Zhang, A., Williamson, D.F., Chen, R.J., Ding, T., Le, L.P., Chuang, Y.-S., Mahmood, F., 2023. Visual language pretrained multiple instance zero-shot transfer for histopathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19764–19775.

Misra, I., Maaten, L.v.d., 2020. Self-supervised learning of pretext-invariant representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6707–6717.

Muljo, H.H., Pardamean, B., Elwirehardja, G.N., Hidayat, A.A., Sudigyo, D., Rahutomo, R., Cenggoro, T.W., 2023. Handling severe data imbalance in chest X-Ray image classification with transfer learning using SwAV self-supervised pre-training. Commun. Math. Biol. Neurosci. 2023, Article–ID.

Noroozi, M., Favaro, P., 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. In: European Conference on Computer Vision. Springer, pp. 69–84.

Oliveira, P., Alliet, M., Coufort-Saudejaud, C., Frances, C., 2018. New insights in morphological analysis for managing activated sludge systems. Water Sci. Technol. 77 (10), 2415–2425.

Ortega, S.T., van den Berg, L., Pronk, M., de Kreuk, M.K., 2022. Hydrolysis capacity of different sized granules in a full-scale aerobic granular sludge (AGS) reactor. Water Res. X 16, 100151.

Ottosen, L.M., Jensen, P.E., Kirkelund, G.M., 2016. Phosphorous recovery from sewage sludge ash suspended in water in a two-compartment electrodialytic cell. Waste Manage. 51, 142–148.

Padilla, R., Netto, S.L., Da Silva, E.A., 2020. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing. IWSSIP, IEEE, pp. 237–242.

Pronk, M., De Kreuk, M., De Bruin, B., Kamminga, P., Kleerebezem, R.v., Van Loosdrecht, M., 2015. Full scale performance of the aerobic granular sludge process for sewage treatment. Water Res. 84, 207–217.

Ranzinger, F., Matern, M., Layer, M., Guthausen, G., Wagner, M., Derlon, N., Horn, H., 2020. Transport and retention of artificial and real wastewater particles inside a bed of settled aerobic granular sludge assessed applying magnetic resonance imaging. Water Res. X 7, 100050.

Remmas, N., Melidis, P., Zerva, I., Kristoffersen, J.B., Nikolaki, S., Tsiamis, G., Ntougias, S., 2017. Dominance of candidate saccharibacteria in a membrane bioreactor treating medium age landfill leachate: Effects of organic load on microbial communities, hydrolytic potential and extracellular polymeric substances. Bioresour. Technol. 238, 48–56.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28, 91–99.

Satoh, H., Kashimoto, Y., Takahashi, N., Tsujimura, T., 2021. Deep learning-based morphology classification of activated sludge flocs in wastewater treatment plants. Environ. Sci.: Water Res. Technol. 7 (2), 298–305.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.

Sorasan, C., Edo, C., González-Pleiter, M., Fernández-Piñas, F., Leganés, F., Rodríguez, A., Rosal, R., 2022. Ageing and fragmentation of marine microplastics. Sci. Total Environ. 827, 154438.

Tijhuis, L., Van Loosdrecht, M., Heijnen, J., 1994. Formation and growth of heterotrophic aerobic biofilms on small suspended particles in airlift reactors. Biotechnol. Bioeng. 44 (5), 595–608.

Wang, Y., Yang, L., Wild, O., Zhang, S., Yang, K., Wang, W., Li, L., 2023. ADMS simulation and influencing factors of bioaerosol diffusion from BRT under different aeration modes in six wastewater treatment plants. Water Res. 231, 119624.

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. https://github.com/facebookresearch/detectron2.

Wu, Y., Ma, X., Guo, G., Jia, T., Huang, Y., Liu, S., Fan, J., Wu, X., 2024. Advancing deep learning-based acoustic leak detection methods towards application for water distribution systems from a data-centric perspective. Water Res. 261, 121999.

Yang, T., Zhang, L., Liu, F., Cheng, C., Li, G., 2024. Hydrodynamic cavitation–impinging stream for enhancing ozone mass transfer and oxidation for wastewater treatment. J. Water Process. Eng. 58, 104799.