



**HAL**  
open science

# Deep learning-based instance segmentation to characterize the morphology of compact aggregates through image analysis

Léo Théodon, Carole Coufort-Saudejaud, Johan Debayle

► **To cite this version:**

Léo Théodon, Carole Coufort-Saudejaud, Johan Debayle. Deep learning-based instance segmentation to characterize the morphology of compact aggregates through image analysis. 14th International Conference on Pattern Recognition Systems (ICPRS-24), Jul 2024, Londres, France. hal-04723423

**HAL Id: hal-04723423**

**<https://hal.science/hal-04723423v1>**

Submitted on 15 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Deep Learning-Based Instance Segmentation to Characterize the Morphology of Compact Aggregates through Image Analysis

Léo Théodon

CNRS, UMR 5307 LGF, Centre SPIN  
MINES Saint-Etienne  
Saint-Etienne, France  
l.theodon@emse.fr

Carole Coufort-Saudejoud

Laboratoire de Génie Chimique  
Université de Toulouse  
CNRS, INPT, UPS, Toulouse, France  
carole.saudejoud@toulouse-inp.fr

Johan Debayle

CNRS, UMR 5307 LGF, Centre SPIN  
MINES Saint-Etienne  
Saint-Etienne, France  
debayle@emse.fr

**Abstract**—The morphological characterization of aggregates is crucial in various industries, affecting the physicochemical properties and functionality of materials. This study develops a dataset of 4,000 synthetic images that closely represent real latex aggregates, validated by Fréchet Inception Distance (FID) computations using a stochastic geometrical model. It also compares five instance-based deep learning segmentation models across three architectures (Mask R-CNN, YOLOv8, and SAM) for analyzing the morphology of latex aggregates. Among them, Mask R-CNN with ResNet101 showed superior segmentation quality. When applied to real images taken at different stages of aggregation, the segmentation results of this model closely matched the experimental observations, demonstrating its capability for detailed morphological analysis.

**Index Terms**—Aggregates, Deep learning, Instance segmentation, Image processing, Neural networks, Object detection

## I. INTRODUCTION

The morphological characterization of aggregates is a critical issue in many scientific fields, from civil engineering [1] to the food [2], chemical, petrochemical and pharmaceutical industries [3, 4]. Indeed, the morphology of aggregates or agglomerates, i.e. their size, shape or texture, has a direct influence on their physico-chemical properties [4] and structure [1], and is therefore directly related to the strength of the materials they constitute, their toxicity, their taste [2] or their impact on health and the environment.

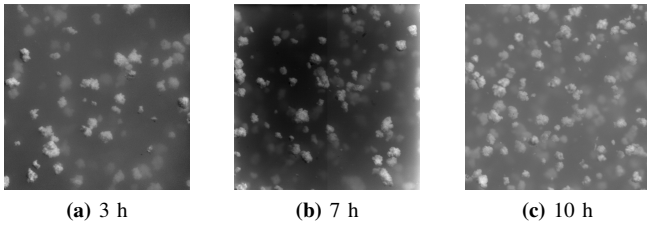
With the increase in computing power and new image processing techniques made possible by the advent of machine learning, the direct measurement of these morphological features through image analysis has seen a surge in popularity in recent years [5]. The use of CNNs [4, 6, 7] or GANs coupled to U-Net type networks [8] for segmentation and characterization of aggregates from electron microscopy images has also gained popularity. In other areas, however, imaging can only be done with optical devices, with a high density of objects, making the image more difficult to process. This may be due to the need to obtain in-situ images at different stages of the aggregation process using non-destructive techniques, or simply because of the size of the objects.

In this case, other approaches to characterize the morphology of the aggregates are possible, such as stochastic methods that rely on models to estimate multiple morphological properties [9]. These approaches are particularly suitable for compact aggregates for which knowledge of 2D morphological features is not sufficient to fully characterize their 3D morphology. However, they also require at least partial knowledge of the 2D morphological properties of the aggregates in order to fine-tune the model parameters. Therefore, this paper proposes to compare different state-of-the-art deep learning instance-based models for compact aggregate segmentation of latex nanoparticle images at different stages of the aggregation process for 2D characterization prior to 3D inference. This is made possible by the generation of realistic and representative images by a stochastic geometrical model, which facilitates the training and comparison of the models.

In the following section, the method of measuring the 2D characteristics of a population of compact aggregates from a series of in-situ images using deep learning instance base segmentation methods is developed. The construction of a training database representative of real data is then detailed. Several deep learning model architectures for instance-based segmentation, such as YOLOv8 [10], Mask R-CNN [11], and SAM [12], are then tested and compared. The results obtained by the best model on real data are presented, and in the final section, improvement possibilities and perspectives are discussed.

## II. METHODOLOGY

The aim of this study is to develop an automatic segmentation method for images of aggregates in dense media. For this purpose, several series of 2,000 images (with a resolution of  $2048 \times 2048$  pixels) were acquired at different stages of the aggregation process of latex nanoparticles in a twelve-liter stirred tank using an optical camera (Fig. 1) [13]. The primary objective is to derive from these image series joint distributions of 2D morphological features, including area, perimeter, circularity, and others, as shown in Table I.



**Fig. 1:** Example of images of the tank taken by the optical device during the aggregation process after 3 hours (a), after 7 hours (b) and after 10 hours (c).

**TABLE I:** List of morphological characteristics

Characteristic	Symbol	Definition and equation
Projected Area	$A$	Area of the object
Convex Area	$A_c$	Area of the convex hull
Perimeter	$P$	Length of the object outline
Equivalent Circle Diameter	ECD	$2 \times \sqrt{A/\pi}$
Convexity	$Co$	$A/A_c$
Circularity	$C$	$4\pi \times A/P^2$

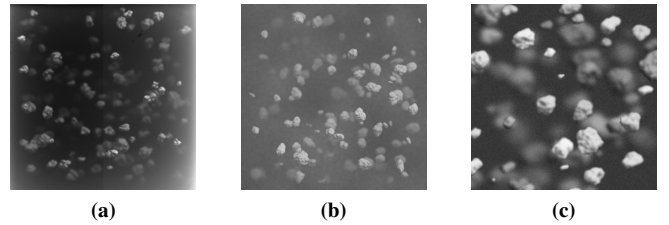
Several methods were considered.

- Direct segmentation methods such as local or adaptive thresholding [14], combined with texture descriptors [15].
- Deep learning semantic segmentation [8, 16].
- Deep learning instance segmentation [4, 7].

The decision to focus on deep learning instance segmentation models was made for several reasons.

- 1) The images captured by the optical device have significant variations in brightness, contrast, and gray balance, making direct thresholding methods inadequate due to their difficulty in generalization.
- 2) Due to the high density of objects, there is a significant amount of overlap, making direct and semantic segmentation methods unsuitable. While these methods are adept at recognizing patterns and textures, the varying distances of objects from the focal plane introduce varying degrees of lens blur, complicating their application.
- 3) Finally, deep learning models, known for their efficiency in real-time image processing, may prove useful in future developments, although not a prerequisite for this study.

Given these considerations, deep learning instance-based segmentation models were selected. Given its prominence in the field of aggregate image segmentation [5], the Mask R-



**Fig. 2:** Examples of synthetic images generated by the SPHERE model.

CNN architecture was evaluated alongside two other state-of-the-art 2023 models: YOLOv8, known for its speed and accuracy, and SAM, a versatile multimodal model.

### III. TRAINING DATABASE CONSTRUCTION

#### A. Database Construction

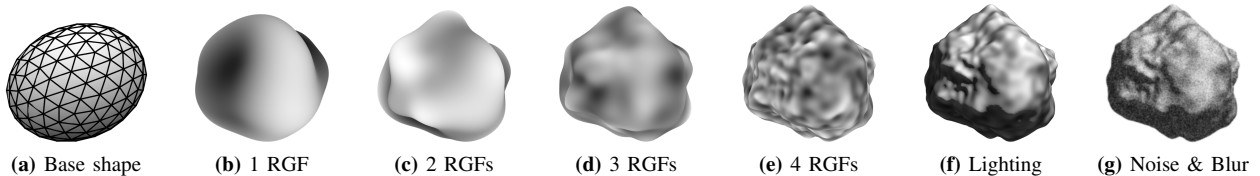
An extended version of the 3D stochastic geometry model known as SPHERE [9] has been used to develop data sets that mimic the real images. The original method, as described by the authors, involves generating granular objects by deforming a mesh with two random Gaussian fields (RGFs). For this research, the method was refined to include four random Gaussian fields to achieve a more realistic rendering. Fig. 3 illustrates the process of creating an aggregate.

Images similar to those shown in Fig. 1 are then generated as follows:

- 1) For each image, a number  $n$  of aggregates are generated using the SPHERE model, where  $n$  varies according to a Poisson distribution parameterized by  $\lambda$ .
- 2) The  $n$  aggregates are uniformly distributed throughout the final image, with successive applications of Gaussian blur to simulate a depth-of-field effect.
- 3) Post-processing techniques such as Gaussian noise and blur are applied to achieve a more realistic rendering.

Examples of synthetic images with different lighting conditions and aggregate concentrations are shown in Fig. 2.

To construct a training dataset that accurately mimics real images, it is crucial to generate synthetic images that closely resemble reality. This involves adjusting the model parameters and 3D rendering settings. The metric used to assess the fidelity of these synthetic images to real images is detailed in the following section.



**Fig. 3:** Sequential generation of an aggregate using an extended version of the SPHERE [9] model: From the base shape (a) to the final rendering (g) with progressive deformations of 1 to 4 random Gaussian fields (b-e), enhanced with lights and shadows (f), and finished with noise and blur effects (g).

## B. Database Validation

To construct the training set, two datasets (Fig. 4) of 2,000 synthetic images each are generated and compared with datasets of 2,000 real images taken 7 hours and 10 hours after the beginning of the aggregation process. The Fréchet Inception Distance (FID), a metric commonly used in GAN-type models [17], is used to evaluate how closely the synthetic images resemble the real ones. The FID calculates the Fréchet distance [18] between high-level features of real and generated images, extracted by a specific layer of the pre-trained Inception-v3 model [19]. A lower FID score indicates a closer similarity between the feature distributions of the generated images and those of the real images, suggesting a higher quality of the synthetic images.

The FID is calculated as follows

- 1) All real and generated images are processed through the Inception-v3 model to extract their features.
- 2) The means  $\mu$  and covariance matrix  $\Sigma$  of the features are calculated for both sets of images.
- 3) The Fréchet distance between the two feature sets is calculated using their means  $\mu_k$  and covariance matrices  $\Sigma_k$ .

The FID formula is given by

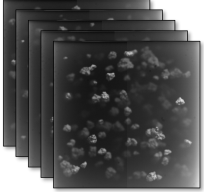
$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr} \left( \Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g} \right) \quad (1)$$

where  $\mu_r$  and  $\mu_g$  are the mean feature vectors of the real and generated images, respectively, and  $\Sigma_r$  and  $\Sigma_g$  are the covariance matrices of the features of the real and generated images, respectively.

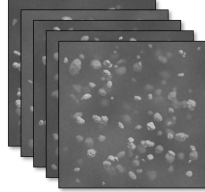
Table II shows the FIDs between datasets A and B, respectively, and the actual images taken at 7 hours (Fig. 1b) and 10 hours (Fig. 1c). The FID scores obtained are both below 100, which is generally considered to be quite good. To assess the quality of the FID scores, the real image datasets at 7 hours and 10 hours were randomly split into subsets of 1000 images. The average FID calculated between these subsets of 1,000 real images at 7 hours and 10 hours was 66.5 and 25.8, respectively, providing a benchmark for good FID performance when comparing real images to synthetic ones.

**TABLE II:** FID score comparison: synthetic datasets versus real images taken at different times

	Dataset A (7 hours)	Dataset B (10 hours)
<b>FID</b>	83.1	45.3

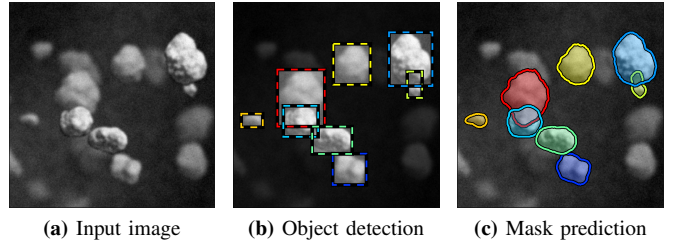


(a) Dataset A,  $\lambda_A = 80$



(b) Dataset B,  $\lambda_B = 130$

**Fig. 4:** Two datasets of 2,000 synthetic images that resemble the real ones are generated.



**Fig. 5:** CNN instance segmentation workflow, from input to object detection to mask prediction.

## IV. SEGMENTATION MODEL COMPARISON

Unlike semantic segmentation, which groups pixels by class without distinguishing between objects, instance segmentation distinguishes individual objects of the same class in an image. Instance segmentation models combine object detection and semantic segmentation through the following simplified steps

- Region Proposal: Identifying potential object locations.
- Bounding Boxes: Outlining detected objects (Fig. 5b).
- Masks: Providing pixel-level object outlines for accurate classification (Fig. 5c).

This approach, exemplified by models such as Mask R-CNN, enables detailed image analysis by identifying and categorizing each object individually.

### A. Model Selection

Several instance segmentation models were evaluated to determine their suitability for the dataset in question. The selection was based on the architecture, performance, and adaptability of each model to the specific task at hand. Mask R-CNN models, proven in the field of instance segmentation of aggregates [6], were selected for comparison with two state-of-the-art architectures introduced in 2023, YOLOv8 and SAM.

Table III summarizes the five architectures evaluated. For mask R-CNN, the ResNet50 and ResNet101 backbones were chosen because the performance improvements of ResNet152 over ResNet101 are relatively small when weighed against the added complexity and computation time [20]. In the case of YOLOv8, the *m* and *x* versions were chosen: *m* offers a balance between performance and accuracy, while *x* is the most sophisticated version (more parameters). SAM was evaluated using only its most advanced version.

**TABLE III:** Comparative overview of tested instance segmentation models

Model	Backbone	Key Features
Mask R-CNN	ResNet50	High accuracy, medium complexity
Mask R-CNN	ResNet101	High accuracy, increased complexity
YOLOv8- <i>m</i>	–	Balance between speed and accuracy
YOLOv8- <i>x</i>	–	High performance, computationally intensive
SAM grand	–	Versatile, capable of segmenting various objects

*Note: "Accuracy" refers to the model's ability to correctly segment and identify objects. "Performance" encompasses both accuracy and the model's processing speed. "Complexity" involves computational resources required and the model's architectural intricacy.*

## B. Training Methodology

The training dataset consists of all images from datasets A and B, with their resolution reduced from  $2048 \times 2048$  pixels to  $1024 \times 1024$  pixels. This reduction aims to reduce training times while maintaining a high level of detail. The dataset of 4000 images is divided into three parts: 3000 images (75%) for training and 500 images (12.5%) each for testing and validation (same proportion from both datasets A and B). To augment the data, the images are rotated in  $90^\circ$  increments, flipped, and scaled by  $\pm 20\%$  of their original size during the training process.

The training process was standardized across models, including both versions of the Mask R-CNN (ResNet50 and ResNet101) and YOLOv8 ( $m$  and  $x$ ) architectures, to ensure comparability. The key parameters are summarized in Table IV, highlighting that the same settings were uniformly applied to maintain consistency in the training conditions.

**TABLE IV:** Summary of training parameters

Parameter	Mask R-CNN	YOLOv8	SAM
Dataset Split	75%/12.5%	75%/12.5%	75%/12.5%
Epochs	50	80	50
Batch Size	32	16	32
Learning Rate	0.001	0.001	0.001

In the case of SAM, a multimodal model, the specification of regions of interest is required. For this purpose, fine grids were generated, with post-processing employed to remove the largest detected objects, such as background elements. Considering a ResNet or YOLO-based model to identify regions of interest for SAM presents an interesting avenue for future research.

Each model, except SAM, was initialized with pre-training weights obtained from training on the COCO dataset [21] to take advantage of its extensive image diversity for improved feature extraction and to ensure equitable training conditions for objective comparison.

## C. Results

To evaluate the performance of the 5 different models, the test set is divided based on the source of the images, distinguishing between Dataset A and Dataset B. The models

are evaluated separately on images from datasets A and B due to significant differences in their generation parameters. The Dataset A has an average of 80 objects per image, a wider histogram range, while the Dataset B has an average of 130 objects, with significantly higher overall brightness and blur (Fig. 6). In addition, the size distribution of the aggregates in the Dataset A follows a beta distribution, while the size distribution of the Dataset B is bimodal, derived from a sum of beta distributions. This distinction makes it possible to assess the ability of the models to accurately identify objects of different sizes within the same image. The usual performance metrics, defined below, are then computed, with the results detailed in Table V.

*Macro Metrics:* For single-class instance segmentation, as it is the case in the context of this study, macro metrics (Accuracy, Precision, Recall, F1 Score) computed from a confusion matrix [22] assess performance by considering each object instance detection as separate, providing a balanced overview of model performance across all instances.

*IoU (Intersection over Union):* A measure of the overlap between predicted and actual object areas, with higher IoUs indicating better segmentation accuracy.

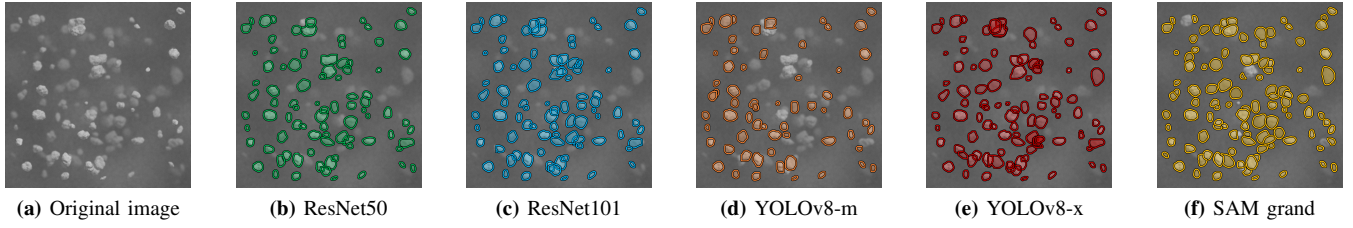
*AP (Average Precision):* As defined in [23], a popular measure for evaluating the balance between precision and recall across various IoU thresholds is calculated in increments of 0.05, ranging from an IoU of 0.5 to 0.95.

*ARE (Absolute Relative Error):* A measure of the accuracy of object counting by comparing the detected object count to the actual object count, with lower ARE indicating higher counting accuracy.

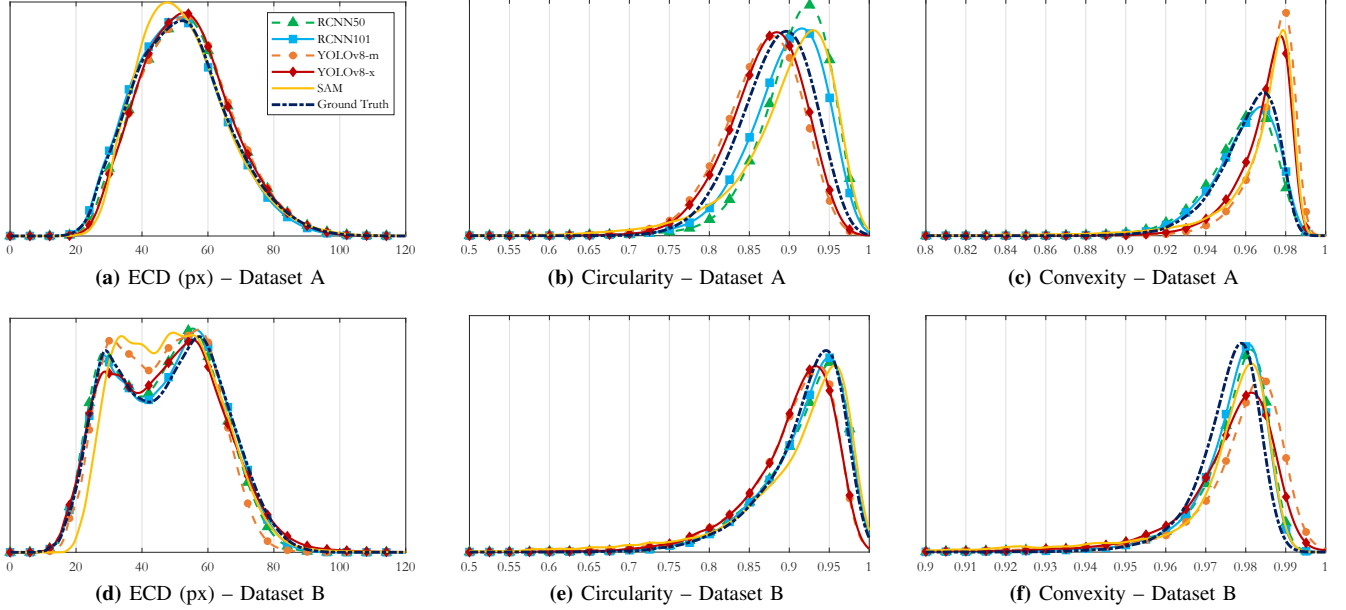
The results presented in Table V show that all models perform better on Dataset A compared to Dataset B. This discrepancy can be attributed to the lower contrast and higher level of blur in Dataset B. In addition, the Absolute Relative Error (ARE) shows a reduced detection rate for Dataset B due to the prevalence of densely packed aggregates and significant blur in the background. The ARE and Fig. 6 also highlights the superior object detection capabilities of the YOLOv8- $x$  and SAM architecture.

**TABLE V:** Comparison of performance metrics across segmentation models for both datasets

Dataset A							
Model	AP	Mean IoU	Accuracy	Precision	Recall	F1 Score	ARE
Mask R-CNN (ResNet50)	0.66	0.801	0.999	0.864	0.888	0.887	12%
Mask R-CNN (ResNet101)	<b>0.67</b>	<b>0.807</b>	0.999	0.900	<b>0.917</b>	<b>0.890</b>	12%
YOLOv8- $m$	0.65	0.796	0.999	0.901	0.873	0.882	10%
YOLOv8- $x$	0.66	0.802	0.999	<b>0.903</b>	0.881	0.886	<b>6%</b>
SAM grand	0.58	0.756	0.998	0.872	0.861	0.850	10%
Dataset B							
Model	AP	Mean IoU	Accuracy	Precision	Recall	F1 Score	ARE
Mask R-CNN (ResNet50)	0.74	0.843	0.999	0.879	0.943	0.912	37%
Mask R-CNN (ResNet101)	<b>0.76</b>	<b>0.853</b>	0.999	0.900	<b>0.953</b>	<b>0.918</b>	33%
YOLOv8- $m$	0.59	0.765	0.999	0.832	0.880	0.860	51%
YOLOv8- $x$	0.66	0.803	0.999	<b>0.904</b>	0.908	0.886	23%
SAM grand	0.65	0.795	0.999	0.882	0.898	0.876	<b>22%</b>



**Fig. 6:** Example of segmentation by the models for an image from dataset B, with recognition rates similar to those shown in Table V (column ARE).



**Fig. 7:** Distributions of size (Equivalent Diameter Circles, EDC) and shape (circularity and convexity) as determined by different models for the test sets of Datasets A and B.

Given the imbalance between object density and background, *accuracy* proves to be an inappropriate metric for this analysis. The Mask R-CNN model using a ResNet101 backbone proves to be the most effective, especially when considering *recall* scores. These scores are crucial as they directly reflect the precision of the predicted segmentation masks.

However, in the context of morphological characterization of aggregates, the standard measures, while essential in deep learning [24], may not be fully sufficient to assess the quality of different models. Therefore, the size and shape distributions of the objects detected by the models are computed and compared to the ground truth (Fig. 7).

The morphological feature distributions show a clear advantage for the Mask R-CNN model with a ResNet101 backbone for both datasets. This difference can be assessed quantitatively by calculating the statistical distance [25] – or Total Variation distance ( $TV$ ) – which is defined as follows

$$TV(p, q) = \frac{1}{2} \int_{-\infty}^{\infty} |p(x) - q(x)| dx \quad (2)$$

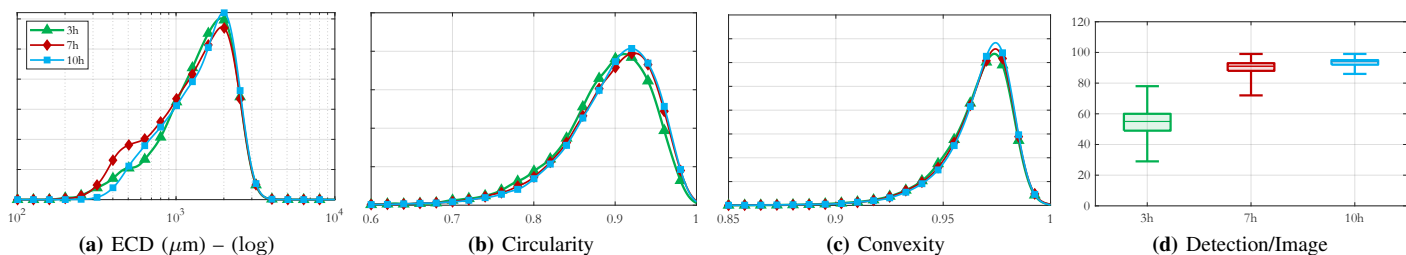
where  $p$  and  $q$  represent the probability density functions of the compared continuous distributions. The Total Variation

**TABLE VI:** Comparison of the Total Variation distance between model-predicted morphological feature distributions and ground truth across datasets A and B

Dataset A					
$TV$	ResNet50	ResNet101	YOLOv8-m	YOLOv8-x	SAM
ECD	0.038	<b>0.017</b>	0.048	0.045	0.045
Circularity	0.230	0.140	0.144	<b>0.101</b>	0.226
Convexity	0.091	<b>0.049</b>	0.390	0.278	0.324
Dataset B					
$TV$	ResNet50	ResNet101	YOLOv8-m	YOLOv8-x	SAM
ECD	0.040	<b>0.016</b>	0.079	0.038	0.101
Circularity	0.045	<b>0.029</b>	0.123	0.126	0.097
Convexity	0.149	<b>0.113</b>	0.295	0.186	0.180

distance ranges from 0 (indicating identical distributions) to 1 (denoting completely distinct distributions).

Table VI confirms that the Mask R-CNN with ResNet101 offers better overall performance than the others. Other measures, such as the Kolmogorov-Smirnov test [26], which determines whether two samples come from the same distribution, also support this result.



**Fig. 8:** Size (ECD) and shape (circularity and convexity) distributions determined by Mask R-CNN with ResNet101, as well as the number of objects detected per image (d) on real image sets acquired at 3 hours, 7 hours, and 10 hours.

#### D. Discussion

When it comes to object detection and size determination (ECD), all models perform quite well, although SAM shows a slight underperformance. Nevertheless, it's important to remember that SAM shows excellent capabilities in terms of object detection ratio, which suggests that its fine tuning could be further optimized. In addition, integration with a YOLOv8 or ResNet-type model for the detection phase could be beneficial and worth evaluating for potential performance gains.

In terms of pure object detection, YOLOv8 models are both faster and more efficient. However, when focusing on the morphological characterization of the objects, and thus the quality of the segmentation, Mask R-CNN outperforms the others in this dataset.

Regarding computational time, both Mask R-CNN and YOLOv8 architectures are well suited for real-time applications. In contrast, the SAM model is significantly more computationally intensive, taking 4 to 5 times longer to process than YOLOv8-x.

#### V. APPLICATION

In this section, the best performing model, Mask R-CNN with ResNet101, is used to segment three sets of images, each consisting of 2,000 images taken at different stages of the aggregation process (3h, 7h, and 10h), as shown in Fig. 1. The results, detailed in Fig. 8, include size and shape distributions and the average number of objects detected per image. Qualitatively, the results are in line with expectations: the number of detectable objects and the proportion of larger objects both increase with time. However, the aggregates stop growing when they reach a critical size, and their shape changes only minimally.

These data are in agreement with experimental observations obtained through direct image analysis and morphogranulometry: the largest objects, which are visible to the camera, tend to adopt spherical and convex shapes [27]. In contrast, the smallest objects, measuring a few hundred microns in size, are generally less circular and convex. However, these smaller objects exceed the detection capabilities of the optical camera, which in this context produced images with a resolution of 1 pixel per 19.58  $\mu\text{m}$ , limiting the ability to see finer details.

#### VI. CONCLUSION

To evaluate the morphological characterization of latex aggregates, the performance of five instance-based segmentation models derived from three different architectures (Mask R-CNN, YOLOv8, and SAM) was compared. A stochastic geometric model generated a database of 4,000 synthetic images for model training. The performance comparison showed that the YOLOv8 and SAM architectures excelled in object detection, while the Mask R-CNN architecture, especially with the ResNet101 backbone, demonstrated superior segmentation quality. This architecture was then used to segment three sets of 2,000 real images acquired at different stages of the aggregation process. The segmentation results were found to be in good agreement with experimental observations.

In terms of future directions, several promising avenues are worth exploring and will be the focus of future work:

- Enriching the database with a more diverse range of object shapes. In addition, incorporating the synthetic image generation process into a more comprehensive workflow, possibly using GANs, could increase the realism and variety of the training data.
- Further refinement of the SAM model for increased accuracy and use of architectures such as YOLOv8 or ResNet to help define regions of interest more efficiently.
- Regarding the morphological characterization of aggregates, the direct prediction of 3D features such as volume by deep learning is a particularly interesting challenge.

This latter aspect will be a priority for future work.

#### ACKNOWLEDGMENT

The author(s) acknowledge(s) the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-20-CE07-0025 (project MORPHING).

#### REFERENCES

- [1] C. Jin, F. Zou, X. Yang, and Z. You, "3d quantification for aggregate morphology using surface discretization based on solid modeling," *Journal of Materials in Civil Engineering*, vol. 31, no. 7, p. 04019123, 2019.
- [2] F. Z. Vissotto, R. C. Giarola, L. C. Jorge, G. T. Makita, G. M. B. Q. Cardozo, M. I. Rodrigues, and F. C. Menegalli, "Morphological characterization with image

- analysis of cocoa beverage powder agglomerated with steam,” *Food Science and Technology*, vol. 34, no. Food Sci. Technol, 2014 34(4), Oct 2014.
- [3] E. M. Alander, M. S. Uusi-Penttilä, and Åke C Rasmuson, “Characterization of paracetamol agglomerates by image analysis and strength measurement,” *Powder Technology*, vol. 130, no. 1, pp. 298–306, 2003.
- [4] P. Monchot, L. Coquelin, K. Guerroudj, N. Feltin, A. Delvallée, L. Crouzier, and N. Fischer, “Deep learning based instance segmentation of titanium dioxide particles in the form of agglomerates in scanning electron microscopy,” *Nanomaterials*, vol. 11, no. 4, 2021.
- [5] L. Théodon, J. Debayle, and C. Coufort-Saudejaud, “Morphological characterization of aggregates and agglomerates by image analysis: A systematic literature review,” *Powder Technology*, vol. 430, p. 119033, 2023.
- [6] M. Frei and F. Kruis, “Image-based size analysis of agglomerated and partially sintered particles via convolutional neural networks,” *Powder Technology*, vol. 360, pp. 324–336, 2020.
- [7] X. Hu, H. Fang, J. Yang, L. Fan, W. Lin, and J. Li, “Online measurement and segmentation algorithm of coarse aggregate based on deep learning and experimental comparison,” *Construction and Building Materials*, vol. 327, p. 127033, 2022.
- [8] B. Rühle, J. F. Krumrey, and V.-D. Hodoroaba, “Workflow towards automated segmentation of agglomerated, non-spherical particles from electron microscopy images using artificial neural networks,” *Scientific Reports*, vol. 11, no. 1, p. 4942, Mar. 2021.
- [9] L. Théodon, C. Coufort-Saudejaud, and J. Debayle, “A stochastic model based on gaussian random fields to characterize the morphology of granular objects,” *Pattern Recognition*, vol. 149, p. 110255, 2024.
- [10] G. Jocher, A. Chaurasia, and J. Qiu, “Ultralytics YOLO,” Jan. 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [12] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” 2023.
- [13] A. Hamieh, C. Coufort-Saudejaud, A. Couffin, A. Liné, and C. Frances, “Morphological properties of latex aggregates generated in agitated reactors,” 2023.
- [14] D. Bradley and G. Roth, “Adaptive thresholding using the integral image,” *Journal of graphics tools*, vol. 12, no. 2, pp. 13–21, 2007.
- [15] J. Chen, G. Zhao, M. Salo, E. Rahtu, and M. Pietikainen, “Automatic dynamic texture segmentation using local descriptors and optical flow,” *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 326–339, 2013.
- [16] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [17] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, G. Klambauer, and S. Hochreiter, “Gans trained by a two time-scale update rule converge to a nash equilibrium,” *CoRR*, vol. abs/1706.08500, 2017. [Online]. Available: <http://arxiv.org/abs/1706.08500>
- [18] D. Dowson and B. Landau, “The fréchet distance between multivariate normal distributions,” *Journal of Multivariate Analysis*, vol. 12, no. 3, pp. 450–455, 1982.
- [19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: <http://arxiv.org/abs/1512.00567>
- [20] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.
- [22] T. Fawcett, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition.
- [23] W. Gu, S. Bai, and L. Kong, “A review on 2d instance segmentation based on deep neural networks,” *Image and Vision Computing*, vol. 120, p. 104401, 2022.
- [24] Y. Tian, D. Su, S. Lauria, and X. Liu, “Recent advances on loss functions in deep learning for computer vision,” *Neurocomputing*, vol. 497, pp. 129–158, 2022.
- [25] A. Tsybakov, *Introduction to Nonparametric Estimation*, ser. Springer Series in Statistics. Springer New York, 2008. [Online]. Available: <https://books.google.fr/books?id=mwB8rUBsbqoC>
- [26] E. Vigneau, C. Loisel, M. Devaux, and P. Cantoni, “Number of particles for the determination of size distribution from microscopic images,” *Powder Technology*, vol. 107, no. 3, pp. 243–250, 2000.
- [27] A. Hamieh, C. Coufort-Saudejaud, A. Couffin, A. Liné, and C. Frances, “Temperature influence on mbs latex aggregate morphology,” *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 676, p. 132139, 2023.