Addressing Data Scarcity in Materials Science Research with Deep Generative Models

Anonymous WACV Applications Track submission

Paper ID 6

Abstract

Developments in deep learning have facilitated the automatic visual analysis of scientific data, driving forward exploratory research. However, these approaches depend on large amounts of expert-annotated data for effective training, which is difficult to come by in narrow application domains. In this work, we address the challenges that come with performing visual analysis of high-speed x-ray phase contrast images of the combustion of molten metal particles. In this case, manual annotations of thousands of complex frames is highly impractical. To address this, we propose a synthetic data generation framework that eliminates the need for large-scale manual labelling by generating image-annotation pairs for the task of image segmentation. We first train a denoising diffusion model with a small number of annotated samples to generate image-binary mask pairs. We use the predictions of a fine-tuned segmentation foundation model to create a multi-class semantic annotations for the synthetic dataset. We apply our framework on x-ray phase contrast videos of particle combustion. From 200 manually annotated frames, we generate 10,000 synthetic image-annotation pairs. We demonstrate that training semantic segmentation models with our generated synthetic data yields significant boost in performance.

1. Introduction

Recent advancements in deep learning have greatly benefited scientific discovery in many fields of research [2, 20]. In particular, this has accelerated the processing of visual information, enabling large scale statistical analysis of data, having applications in fields such as cell microscopy [7, 24], bioimage analysis [9], and high energy physics [5]. However, these data-driven approaches depend on the availability of large amounts of data annotated for the required visual perception task. This becomes particularly challenging when applying general deep learning tools in narrow domain scenarios such as those listed above. This is due to the fact that domain-specific, expert-level knowledge is required for annotating images, and obtaining a large amount of annotated data becomes expensive and impractical. There are a few recent studies proposed synthetic data generation pipelines. In [13], a framework for synthetic data generation using large language models and text-to-image diffusion models was proposed, employing cross-attention maps to generate pseudo-semantic segmentation labels. Similarly, [22] transforms diffusion models into tools for generating labeled datasets, introducing a perception decoder to produce high-quality labels like segmentation masks and depth maps alongside synthetic samples. In [4], authors proposed a framework to generate pseudo binary segmentation labels utilizing off-the-shelf medical foundation models.

In this work, we address this challenge as it applies to materials science research. Specifically, we look at a large dataset of high-speed x-ray phase contrast imaging (XPCI) videos [21], that captures in-situ behavior of molten metal particles as they combust. Elemental and composite metal powders are useful for combustion applications due to their high combustion enthalpies and burn temperatures [15]. This dataset characterizes the combustion properties of reactive metal powders that could be used in munitions for rapid neutralization of chemical warfare agents (CWAs). Predicting particle attributes such as behavior, shape, trajectory, and frequency of occurrence thus becomes an important task for research of these processes. However, due to the complexity of the combustion scenes tracking the evolution of particles' combustion behavior cannot be done without using sophisticated and high-throughput data driven methods.

In order to leverage deep learning tools for visual reasoning tasks in the absence of a large-scale annotated dataset, we propose a framework for the generation of synthetic XPCI images of particle combustion events along with their semantic segmentation annotation masks. Specifically, we train a denoising diffusion probabilistic model (DDPM) [11] on a small number of manually annotated images and generate a large dataset of image and binary mask pairs.

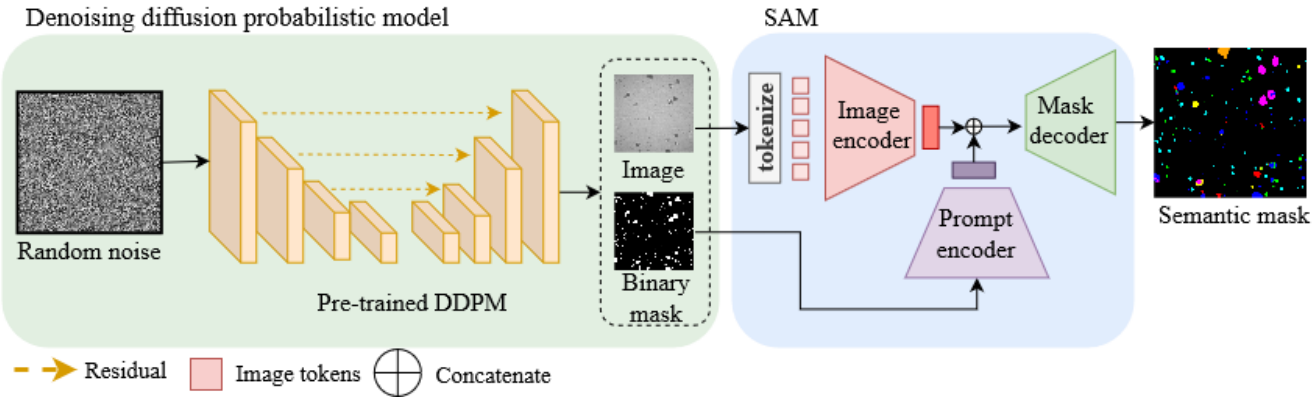

Figure 1. An illustration of the synthetic dataset creation pipeline using a pre-trained diffusion model and the fine-tuned SAM. In this framework, we employ a DDPM to synthesize realistic images together with their corresponding binary masks, which are then fed into the fine-tuned SAM model; SAM uses these binary masks as prompts to generate pseudo-annotated synthetic images with detailed semantic segmentation masks.

Here, the binary mask indicates the location of all types of particles present in the image. We then leverage the foundation Segment Anything Model (SAM) [12] fine-tuned on the same small set of manually annotated images to predict semantic categories of each particle, using the binary mask as a prompt to SAM. Through this process, we create a large dataset of synthetic images and their corresponding semantic pseudo-annotations. We demonstrate the effectiveness of the additional synthetic training data by training state-of-the-art segmentation baselines [3, 10, 16]. We show that training these models on a combination of real and synthetic images results in a significant boost in performance.

- We propose a framework for the generation of realistic images-mask pairs of x-ray phase-contrast images of particle combustion events to address the scarcity of annotated data for the task of semantic segmentation.
- With the addition of our generated synthetic data in the training set, we demonstrate boosts in segmentation performance of TransUNet [3], SwinUNETR [10] and EMCAD [16] across multiple metrics.

2. Methodology

The objective of this work is to generate a large number of high-quality synthetic images of particle combustion events and their corresponding semantic segmentation masks. An illustration of the framework can be seen in Figure 1.

First, a diffusion-based generation model is trained on a small number of manually annotated frames to generate realistic synthetic images along with their binary segmentation masks. In order to achieve realistic binary mask generation via learning joint distribution, we concatenate target images with the corresponding ground truth binary masks and define the diffusion objective with this concatenated representation. This process is detailed in Section 2.1. Through this step, we obtain synthetic frames of particle combustion events and their corresponding binary masks showing particle locations without class labels.

Next, we assign semantic categories to each particle mask generated in the previous step. Towards this, we adapt SAM with the same small set of annotated real images. This step enhances the model's capability to accurately annotate the synthetic images, creating reliable pseudo-annotations that closely mimic expert annotations. Figure 1 illustrates the overall framework, from synthetic image and binary mask generation to leveraging adapted SAM to predict semantic categories. Figure 2 illustrates the SAM adaptation framework.

Finally, we train a segmentation network with the synthetic dataset generated in the previous steps. This strategy compensates for the limited availability of real annotated images in materials science, ultimately improving the segmentation performance. Each step of the framework is detailed below.

2.1. Synthetic image generation

We utilize denoising diffusion probabilistic models (DDPMs) for image generation due to their demonstrated ability in learning complex data distributions and creating high quality synthetic images [11, 14]. DDPM relies on two distinct phases, forward and backward diffusion. In the forward diffusion process, normally distributed random noise is gradually added to the data sample with a predetermined variance schedule until time index T, then during backward diffusion phase, a deep denoiser network, typically a UNet [11], is employed to predict the added noise in each time step. The forward diffusion function can be described as

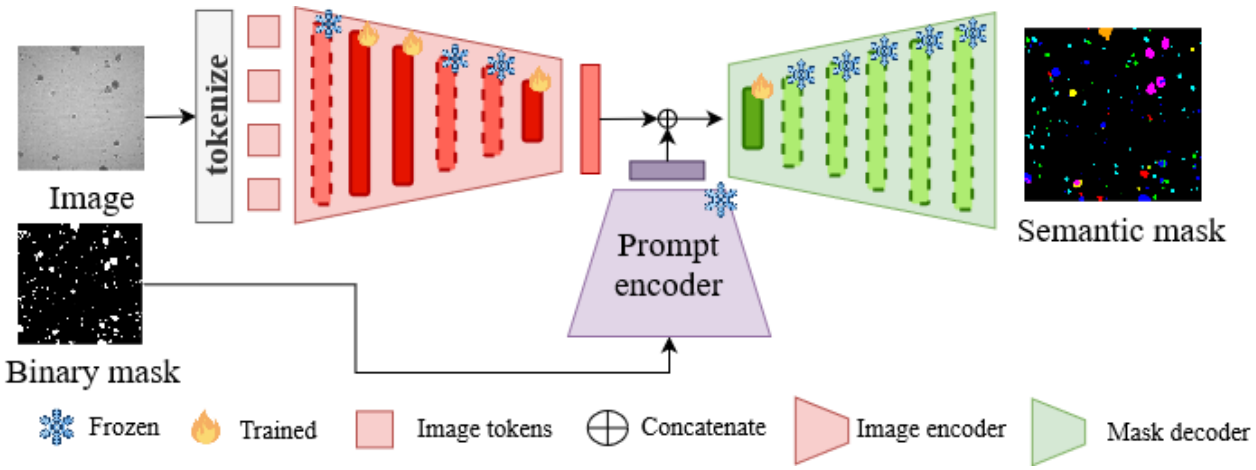

Figure 2. An illustrative overview of the proposed SAM fine-tuning framework, highlighting which components of the architecture are updated during training and which remain frozen. For each block in the network, we indicate its optimization status using visual markers: layers annotated with a fire icon are trainable, whereas those marked with a cold icon are kept fixed. In this setup, we first tokenize the input images and feed them to the Vision Transformer encoder, while binary masks indicating the locations of particles in the image are provided to the prompt encoder; the mask decoder then combines these representations to produce the final semantic segmentation masks.

$$x_t = \sqrt{\alpha_t}\, x_{t-1} + \sqrt{1-\alpha_t}\, \epsilon \tag{1}$$

where α_t denotes the pre-fixed variance schedule, x_t and x_{t-1} are the data samples at time t and $t-1$ respectively and $\epsilon \sim \mathcal{N}(0, I)$. After the forward diffusion process, a deep denoiser network is trained to learn the mapping between data samples from time t to $t-1$ via estimating added noise, the objective function can be written as follows

$$L(\theta) = \mathbb{E}_{t,\mathbf{x}_0,\epsilon}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t\right)\right\|^2\right], \tag{2}$$

where $\bar{\alpha}_t = \prod_{m=1}^{t} \alpha_m$, x_0 is the data sample at time zero and $\boldsymbol{\epsilon}_\theta(.)$ denotes the predicted noise with network parameters θ. In our pipeline, we train an unconditional improved-DDPM [14] using the available small annotated set and their corresponding binary masks. Therefore, with the above notation x_0 becomes

$$x_0 \in \mathbb{R}^{H\times W\times 2}, \quad x_0 = \left(x_0^{(1)}, x_0^{(2)}\right), \tag{3}$$

where $x_0^{(1)} \in \mathbb{R}^{H\times W}$ and $x_0^{(2)} \in \mathbb{R}^{H\times W}$ denote clean image and ground truth binary mask with height (H) and width (W) respectively. We generate binary masks to be used in training from the annotated semantic masks. Then, during the inference we synthesize images and corresponding binary masks from random noise with improved-DDPM [14] sub-sampling algorithm. In our experiments we observed the binary mask generation along with the synthetic image is beneficial to guide SAM during the fine-tuning phase by directing the network through prompt encoder (see Figure 2 for details). A few synthesized samples generated with our diffusion model along with real samples are illustrated in Figure 5 and Figure 4.

2.2. SAM Adaptation

The Segment Anything Model (SAM) is composed of three sub-networks: an image encoder, a prompt encoder, and a mask decoder. The image encoder is a vision transformer [6] trained to extract meaningful features from the input image. The prompt encoder leverages convolutional layers to incorporate informative prompts, such as points, bounding boxes, or image masks, enhancing the accuracy of object segmentation. Finally, the mask decoder combines the encoded image and prompt vectors using a transformer layer and then utilizes transposed convolutions to up-sample and predict precise segmentation masks.

Following [8], we incorporate small adapter networks consisting of multi-layer perceptrons into the attention blocks of transformer layers of both the image encoder and mask decoder networks. During training, only the adapter layers are updated. This decreases the number of trainable parameters and allows the network to retain the knowledge present in the original weights while also fine-tuning the model for our specific task. We utilize a small set of annotated samples and their corresponding binary masks with a combination of cross-entropy and dice losses [18] during fine-tuning. Throughout the process, we constantly used binary masks as the prompts apart from the target images to guide the network more precisely to deal with the complex nature of X-ray phase contrast images and small particle sizes. The fine-tuning procedure is illustrated in Figure 2. We illustrate the quality of the pseudo-annotations per-

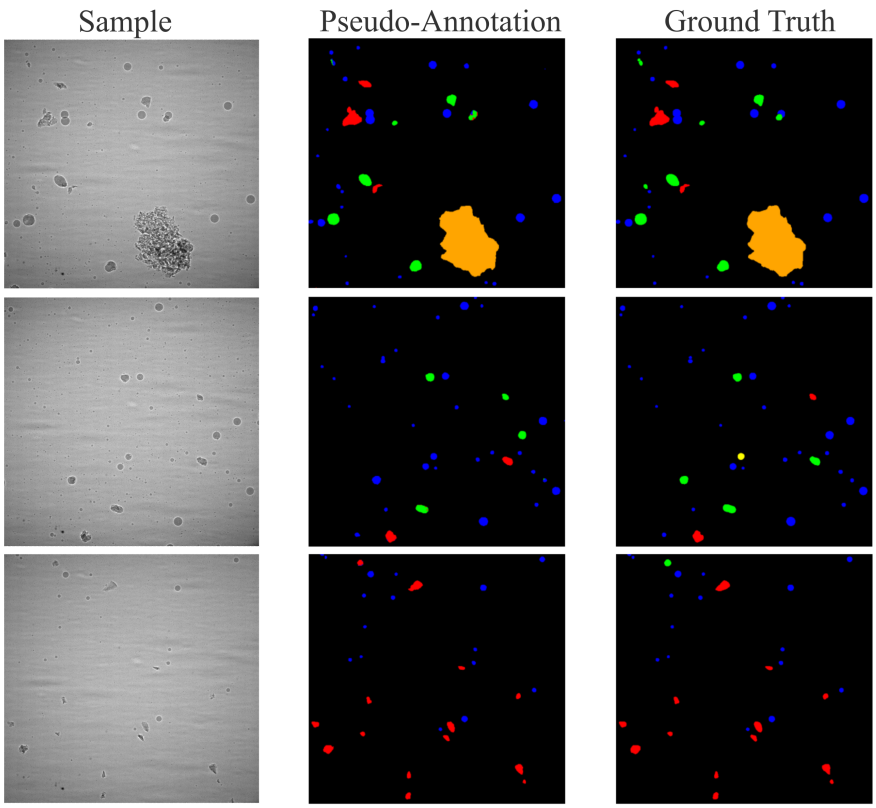

Figure 3. Generated pseudo-annotations for real XPCI images, all produced by our fine-tuned SAM model, compared against the corresponding ground truth segmentation masks. The qualitative alignment between the pseudo-labels and the manually annotated masks highlights the accuracy and reliability of our pseudo-annotation pipeline, indicating that the fine-tuned SAM can effectively capture particle locations and shapes in realistic experimental data.

formed by fine-tuned SAM in Figure 3.

2.3. Semantic segmentation

We combined real and synthetic images in varying proportions, progressively increasing the number of synthetic images in the final training dataset. While our framework is compatible with any segmentation network, we specifically selected three state-of-the-art architectures—TransUNet [3], a vision transformer-based model; SwinUNETR [10], a Swin transformer-based network; and EMCAD [16], a CNN-based model—to demonstrate its efficiency across different backbone designs.

3. Experiments

We performed an extensive evaluation to show the benefits of our proposed approach in generating synthetic datasets. Our results are grouped with different proportion of the synthetic samples in the final training set. We employ the Dice similarity coefficient [18], normalized surface Dice (NSD) [17], and micro F1 scores as the primary metrics for comparison in our evaluation. The Dice and NSD scores are used to assess the overall overlap between the target and predicted segmentation masks, providing an overall measure of segmentation quality. The micro F1 score accounts for the occurrence of false positives and false negatives for all classes, considering the imbalanced nature of the dataset, contributing to a more robust evaluation.

XPCI Dataset: High-speed XPCI is performed on a variety of aluminium-based metal composite and pure titanium, hafnium, and zirconium powders using a high energy x-ray synchrotron to assess how variations in chemistry impact combustion behavior. These experiments were performed in a similar manner with a single Al:Zr powder chemistry [19] and utilize synchronous hyper-spectral imaging of the combustion scene to measure combustion temperatures [1]. This resulted in the capture of a data set of 500,00 video frames containing tens of thousands of particles creating a highly imbalanced segmentation dataset due to small particle sizes compared to the background, and also due to underrepresented particle classes. We selected a 200 sample subset of the XPCI for manual annotation each with size 624×640

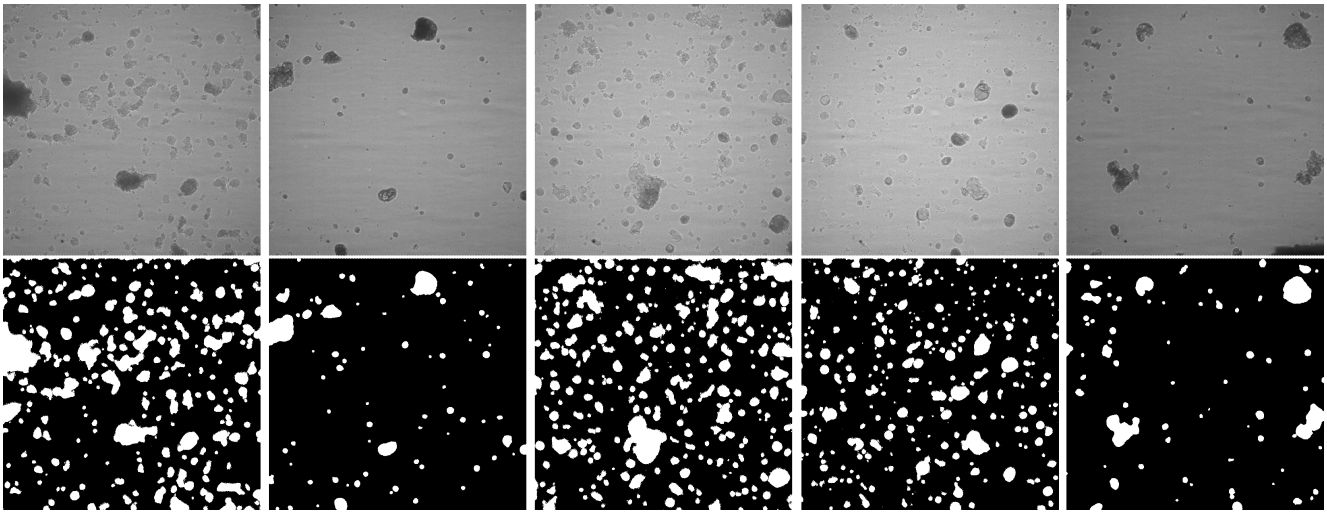

Figure 4. Visualization of synthetic images and their corresponding binary masks, all generated automatically by our proposed pipeline. The examples illustrate the diversity and realism of the synthesized particle patterns, as well as the precise spatial alignment between the images and their binary annotations.

Table 1. Semantic segmentation performance of TransUNet [3] with different numbers of synthetic images generated with SAM-t.

Dataset	Dice Score	NSD	Micro F1
200 real images	0.7791	0.5490	0.6933
200 real images + 500 synthetic images	0.7807	0.5195	0.6854
200 real images + 1,000 synthetic images	0.7985	0.5731	**0.7611**
200 real images + 2,000 synthetic images	0.7936	0.5605	0.7508
200 real images + 5,000 synthetic images	**0.8034**	**0.5780**	0.7531

Table 2. Semantic segmentation performance of Swin UNETR [10] with different numbers of synthetic images generated with SAM-b.

Dataset	Dice Score	NSD	Micro F1
200 real images	0.7628	0.5142	0.7243
200 real images + 500 synthetic images	0.7646	0.5233	0.7373
200 real images + 1,000 synthetic images	0.7755	0.5127	0.7272
200 real images + 2,000 synthetic images	0.7756	0.5393	0.7444
200 real images + 5,000 synthetic images	**0.7884**	**0.5576**	**0.7531**

pixels (shown in Figure 5), featuring 9 distinct classes of particles derived from X-ray phase contrast videos of combusting composite and elemental powders. Additionally, we utilized a separate subset of 59 annotated frames as the test set to evaluate the final segmentation performance.

3.1. Implementation details

We utilized a workstation equipped with 8 NVIDIA RTX A5000 GPUs with 24GB memory for all experiments. During the pre-training of diffusion model, we utilize the improved-DDPM [14] pipeline with cosine noise schedule, 4000 forward diffusion steps, 250 sampling steps and AdamW optimizer with a learning rate of 10^{-4}. DDPM training is lasted for 50,000 steps with a batch size of 1.

We utilize two variants of SAM, SAM-base (SAM-b) [12] and SAM-tiny (SAM-t) [23] where SAM-tiny utilizes a much smaller image encoder trained via knowledge distillation, while SAM-base utilizes a larger image encoder (ViT-B [6]). During the SAM fine-tuning and final model training, AdamW optimizer is utilized with beta parameters (0.99, 0.999) and a weight decay factor of 0.1. SAM is fine-tuned for 100 epochs with batch size of 1 and 10^{-4} learning rate with polynomial decay rate with power 0.9. Similarly, TransUNet, SwinUNETR and EMCAD are trained for 100 epochs with learning rate 10^{-4} with a batch size of 4.

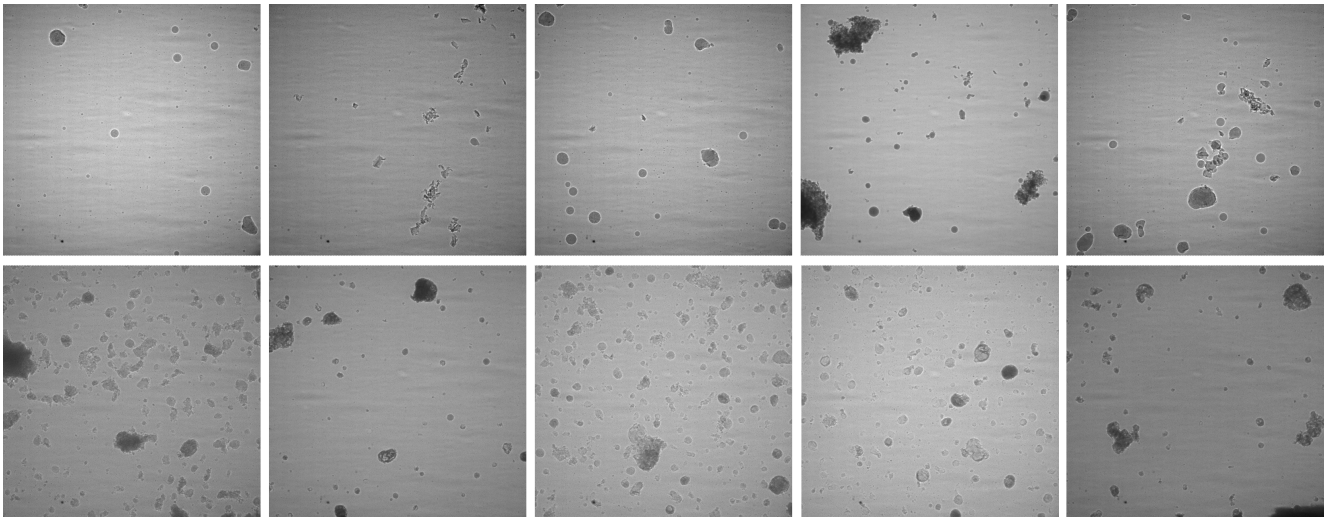

Figure 5. Visualization of real samples from the XPCI particle combustion dataset (first row) and generated synthetic images (second row). Please zoom-in for the best quality.

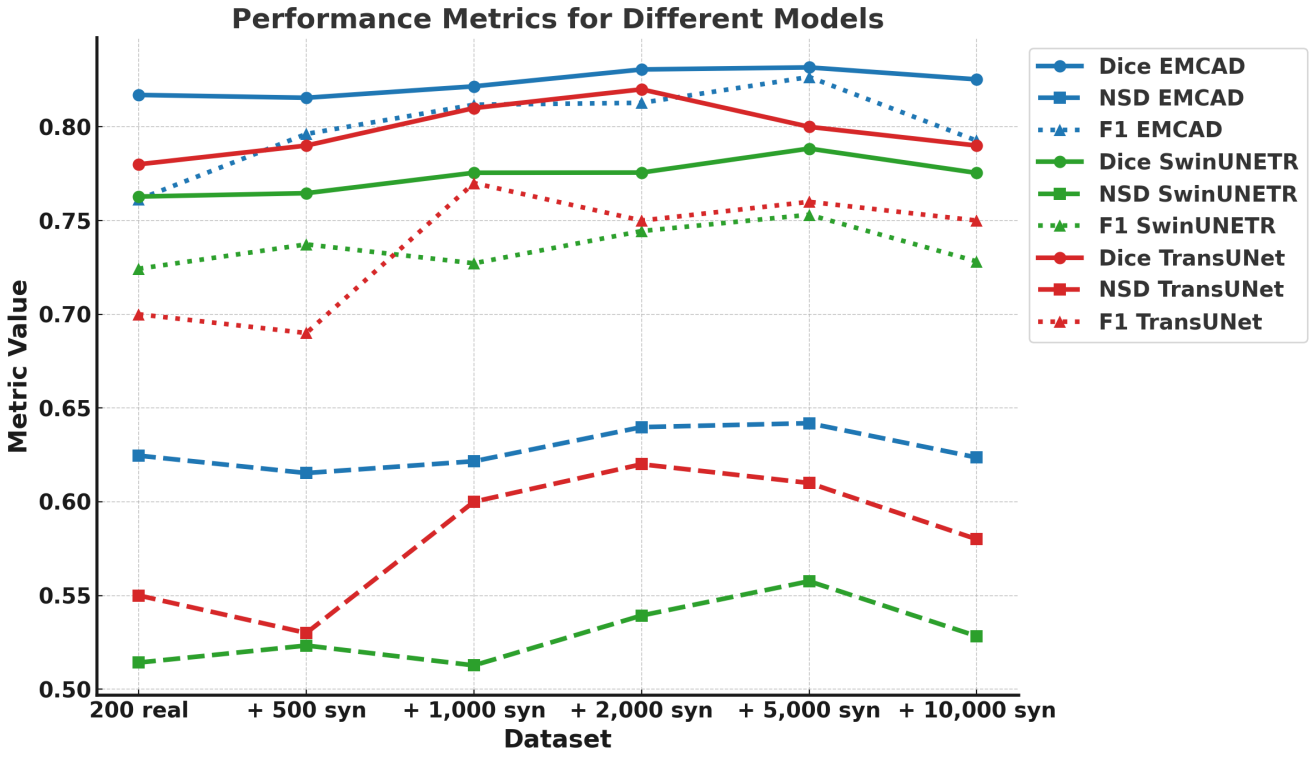

Figure 6. Semantic segmentation performance of different backbones against increasing amounts of synthetic training data.

3.2. Results

We demonstrate the effectiveness of our dataset creation pipeline by measuring the final segmentation performance under different proportions of synthetic samples in the training dataset, starting from a small only real dataset (with 200 samples) and gradually increasing the synthetic sample size.

We illustrate the performance metrics of different segmentation models as a graph in Figure 6. From the graph, we observe that adding 1,000 to 5,000 synthetic images provides a noticeable boost in performance, suggesting that a moderate amount of synthetic data can enhance segmentation performance by providing additional training diversity.

Table 3. Semantic segmentation performance of EMCAD [16] with different numbers of synthetic images generated with SAM-b.

Dataset	Dice Score	NSD	Micro F1
200 real images	0.8170	0.6246	0.7613
200 real images + 500 synthetic images	0.8155	0.6153	0.7962
200 real images + 1,000 synthetic images	0.8216	0.6216	0.8118
200 real images + 2,000 synthetic images	0.8306	0.6398	0.8128
200 real images + 5,000 synthetic images	**0.8317**	**0.6419**	**0.8266**

Table 4. Semantic segmentation performance of TransUNet [3] with different numbers of synthetic images generated with SAM-b.

Dataset	Dice Score	NSD	Micro F1
200 real images	0.7791	0.5491	0.6933
200 real images + 500 synthetic images	0.7807	0.5387	0.6993
200 real images + 1,000 synthetic images	0.8100	0.6034	**0.7752**
200 real images + 2,000 synthetic images	**0.8196**	**0.6126**	0.7504
200 real images + 5,000 synthetic images	0.8058	0.5963	0.7656

Beyond 5,000 synthetic images, the improvements plateau or slightly decline, indicating diminishing returns. This suggests that while synthetic data is beneficial up to a point, too much synthetic data could lead to overfitting or redundancy for these models.

Detailed metrics with synthetic dataset generated with SAM-b have been shown in Tables 4, 3, and 2 for TransUNet, EMCAD and SwinUNETR respectively. From the tables, we observe that the micro-F1 score shows a greater improvement with the addition of synthetic samples compared to the NSD and Dice scores, suggesting that the model improves with synthetic samples at managing the imbalanced dataset by reducing overall false positives and false negatives across all classes.

Ablation Studies: We analyze the effects of different network components on the final segmentation performance. First, we evaluate the quality of synthetic datasets generated using different variants of SAM, specifically SAM-t and SAM-b shown in Tables 1 and 4 respectively, which correspond to different model sizes. In Table 4, there is a noticeable improvement in performance metrics compared to the baseline up to 2,000 synthetic samples, where the highest performance is observed. However, Table 1 reveals a different pattern, with peak performance achieved at 5,000 synthetic samples. This suggests that the quality of the synthetic dataset significantly impacts the performance metrics, implying that lower-quality synthetic datasets require a larger number of samples to achieve comparable performance. Second, we demonstrated the impact of binary mask prompts on the segmentation performance of SAM, highlighting their role in enhancing the quality of the synthetic datasets. Table 5 demonstrates that the performance gain from using binary mask prompts is much more pronounced for smaller-sized model. This finding highlights that smaller SAM variants, like SAM-t, benefit significantly from the inclusion of binary masks during synthetic dataset generation likely due to their limited capacity to capture complex features without additional guidance.

Table 5. Semantic segmentation performance of fine-tuned SAM with different configurations.

Model & Configuration	Dice Score
SAM-t without binary mask prompt	0.6610
SAM-t with binary mask prompt	0.7876
SAM-b without binary mask prompt	0.8127
SAM-b with binary mask prompt	0.8181

4. Conclusion

In this work, we addressed the critical challenge of annotated data scarcity in materials science research, with a particular focus on the localization and categorization of particles in XPCI images of combustion events. To mitigate the high cost and difficulty of obtaining expert-labeled data in this domain, we proposed a complete pipeline for generating a large dataset of synthetic images together with their corresponding semantic segmentation annotations. Our approach leverages synthetic data to approximate the visual characteristics and particle distributions observed in real experimental XPCI frames, while automatically producing pixel-level labels that would otherwise require extensive manual effort.

We evaluated the effectiveness of the proposed pipeline

by training state-of-the-art semantic segmentation models on different mixtures of real and synthetic data, systematically varying the proportion of synthetic samples used during training. The experimental results show that models trained with the inclusion of our generated synthetic data consistently achieve improved performance compared to those trained solely on limited real data, indicating that the synthetic dataset provides meaningful and complementary supervision. These findings suggest that our synthetic data generation framework can serve as a practical and scalable solution for enhancing model performance in scenarios where high-quality annotations are scarce or expensive to obtain.

References

[1] M. Alemohammad, E. R. Wainwright, J. P. Stroud, T. P. Weihs, and M. A. Foster. Kilohertz frame rate snapshot hyperspectral imaging of metal reactive materials. *Applied Optics*, 59, 2020. 4

[2] R. Chang, Y.-X. Wang, and E. Ertekin. Towards overcoming data scarcity in materials science: unifying models and datasets with a mixture of experts framework. *npj Computational Materials*, 8(1):242, 2022. 1

[3] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 2, 4, 5, 7

[4] T. Deshpande, E. Prakash, E. G. Ross, C. Langlotz, A. Ng, and J. M. J. Valanarasu. Auto-generating weak labels for real & synthetic data to improve label-scarce medical image segmentation. *arXiv preprint arXiv:2404.17033*, 2024. 1

[5] F. A. Di Bello, S. Ganguly, E. Gross, M. Kado, M. Pitt, L. Santi, and J. Shlomi. Towards a computer vision particle flow. *The European Physical Journal C*, 81:1–14, 2021. 1

[6] A. Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3, 5

[7] M. Eisenstein. Ai under the microscope: the algorithms powering the search for cells. *Nature*, 623(7989):1095–1097, 2023. 1

[8] H. Gu, H. Dong, J. Yang, and M. A. Mazurowski. How to build the best medical image segmentation algorithm using foundation models: a comprehensive empirical study with segment anything model. *arXiv preprint arXiv:2404.09957*, 2024. 3

[9] A. Hallou, H. G. Yevick, B. Dumitrascu, and V. Uhlmann. Deep learning for bioimage analysis in developmental biology. *Development*, 148(18):dev199616, 2021. 1

[10] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer, 2021. 2, 4, 5

[11] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1, 2

[12] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 2, 5

[13] Q. Nguyen, T. Vu, A. Tran, and K. Nguyen. Dataset diffusion: Diffusion-based synthetic data generation for pixel-level semantic segmentation. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[14] A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. 2, 3, 5

[15] S. M. Peiris and N. Bolden-Frazier. Applications of reactive materials in munitions. In *Shock Phenomena in Granular and Porous Materials*, pages 165–191. Springer, 2019. 1

[16] M. M. Rahman, M. Munir, and R. Marculescu. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11769–11779, 2024. 2, 4, 7

[17] S. Seidlitz, J. Sellner, J. Odenthal, B. Özdemir, A. Studier-Fischer, S. Knödler, L. Ayala, T. J. Adler, H. G. Kenngott, M. Tizabi, M. Wagnet, F. Nickel, B. Müller-Stich, and L. Maier-Hain. Robust deep learning-based semantic organ segmentation in hyperspectral images. *Medical Image Analysis*, 80: 102488, 2022. 4

[18] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, MLCDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017. 3, 4

[19] E. R. Wainwright, S. V. Lakshman, A. F. T. Leong, A. H. Kinsey, J. D. Gibbins, S. Q. Arlington, T. Sun, K. Fezzaa, T. C. Hufnagel, and T. P. Weihs. Viewing internal bubbling and microexplosions in combusting metal particles via x-ray phase contrast imaging. *Combustion and Flame*, 199:194–203, 2019. 4

[20] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T.-Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C. W. Coley, Y. Bengio, and M. Zitnik. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023. 1

[21] S.W. Wilkins, T.E. Gureyev, D. Gao, and A.W. Stevenson. Phase-contrast imaging using polychromatic hard x-rays. *Nature*, 384:335–338, 1996. 1

[22] W. Wu, Y. Zhao, H. Chen, Y. Gu, R. Zhao, Y. He, H. Zhou, M. Z. Shou, and C. Shen. Datasetdm: Synthesizing data with perception annotations using diffusion models. *Advances in Neural Information Processing Systems*, 36:54683–54695, 2023. 1

[23] C. Zhang, D. Han, Y. Qiao, J. U. Kim, S.-H. Bae, S. Lee, and C. S. Hong. Faster segment anything: Towards lightweight sam for mobile applications. *arXiv preprint arXiv:2306.14289*, 2023. 5

[24] S. Zhou, B. Chen, E. S. Fu, and H. Yan. Computer vision meets microfluidics: a label-free method for high-throughput cell analysis. *Microsystems & Nanoengineering*, 9(1):116, 2023. 1