

# IPSM-Bench: A New Intermediate Phase Segmentation Benchmark in Microstructure Images of Zinc-Based Absorbable Biomaterials

Jinglin Xu<sup>1</sup>, Shangyan Zhao<sup>4</sup>, Jiabo Wang<sup>1</sup>, Xinghong Mu<sup>1</sup>, Yulong Lei<sup>1</sup>, Jiacheng Zhang<sup>6</sup>, Hongbo Sun<sup>3\*</sup> and Yageng Li<sup>2,5\*</sup>

<sup>1</sup>School of Artificial Intelligence, University of Science and Technology Beijing, China

<sup>2</sup>School of Advanced Materials Innovation, University of Science and Technology Beijing, China

<sup>3</sup>China Telecom Artificial Intelligence Technology (Beijing) Co., Ltd

<sup>4</sup>School of Materials Science and Engineering, University of Science and Technology Beijing, China

<sup>5</sup>Institute of Materials Intelligent Technology, Liaoning Academy of Materials, China

<sup>6</sup>School of Big Data and Software Engineering, Chongqing University, China

## Abstract

Zinc-based alloys are indispensable emerging absorbable metallic biomaterials, and their macroscopic performance is governed by microstructural characteristics. Intermediate phases—key microstructural constituents—are pivotal in regulating mechanical and functional properties. However, intermediate phase segmentation in zinc alloy microstructures faces formidable challenges: scarce annotated datasets, low contrast, difficulty detecting small targets, and heterogeneous morphologies. To this end, we construct IPSM-Bench, the largest high-quality dataset for zinc-alloy intermediate phase segmentation. Furthermore, we propose SCoP-SAM, a new Spatial Context Prior-guided SAM method that leverages the gradient structure and grayscale properties of intermediate phases to capture spatial context priors and incorporates them into the entire SAM encoding-decoding process, improving segmentation performance. Based on the proposed IPSM-Bench, we establish a new benchmark for intermediate phase segmentation to systematically evaluate state-of-the-art (SOTA) methods and advance research on zinc alloy microstructure analysis. Extensive experiments on IPSM-Bench and additional public alloy benchmarks demonstrate that our SCoP-SAM not only achieves SOTA performance for zinc-alloy intermediate phase segmentation but also generalizes remarkably well to other alloy scenarios.

## 1 Introduction

Zinc-based alloys (i.e., zinc alloys), an indispensable class of emerging and promising absorbable metallic biomaterials, have substantial application potential in orthopedics [Huang *et al.*, 2025], cardiovascular medicine [Yang *et al.*, 2023], and dentistry [Tong *et al.*, 2023]. The microstructural characteristics of zinc alloys directly dictate their macroscopic perfor-

mance, and intermediate phases—key microstructural components—play a pivotal role in regulating mechanical and functional properties. In real-world scenarios, the microstructure of zinc alloys exhibits complex multiphase characteristics; for example, Zn-Mg alloys comprise both lamellar eutectic morphologies and spherical precipitates. These intermediate phases span a broad size spectrum, exhibit uneven spatial dispersion, and display heterogeneous morphologies (e.g., lamellar, spherical, dendritic).

Traditional methods for representing intermediate phases primarily rely on techniques such as scanning electron microscopy (SEM) and optical microscopy (OM) to acquire microstructural images, followed by segmentation and statistical analysis utilizing manual annotations by materials professionals and simple image analysis software. For one thing, manual annotation is inefficient and fails to meet the stringent requirements for accuracy and reproducibility. For another, existing image segmentation techniques face many challenges in zinc alloy microstructural analysis: 1) **Scarcity of Annotated Datasets**: High-quality microstructural images of zinc alloys incur substantial acquisition costs and require precise annotation by materials science experts, resulting in a severe paucity of available training data. 2) **Low Contrast**: The grayscale contrast between the matrix and intermediate phases in zinc alloys is often marginal, especially in SEM backscattered electron images, where the contrast between phases can differ by only a few grayscale levels. 3) **Difficulty in Small Target Detection**: Zinc alloys contain a large number of submicron-scale intermediate phases, which occupy an extremely small pixel fraction in images, carry negligible semantic and textural information, and are highly susceptible to confusion with background textures. 4) **Challenges in Complex Morphologies**: Intermediate phases in zinc alloys exhibit complex, heterogeneous morphologies, ranging from irregular forms (e.g., lamellar, spherical, dendritic, needle-like), often accompanied by overlap and agglomeration. In summary, mitigating the scarcity of annotated datasets and accurately segmenting intermediate phases are critical prerequisites for addressing these challenges.

Therefore, we first construct a large-scale, high-quality annotated dataset of microstructural images of zinc alloys,

\*Corresponding Authors

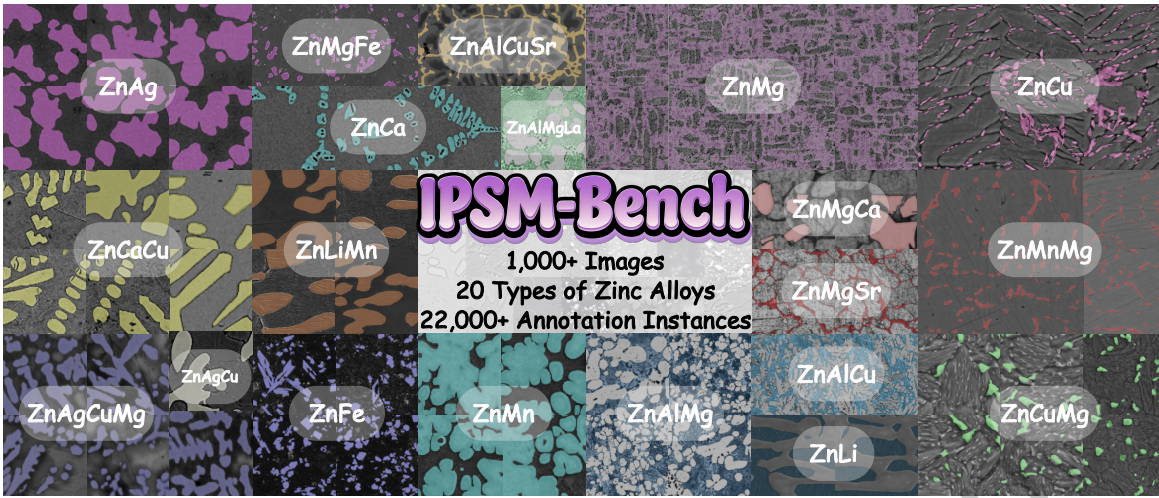


Figure 1: Overview of the proposed IPSM-Bench.

| Dataset                                | Material                         | Available  | Anno.Type   | Resolution     | Alloy Types | Number of Images | Data Modality |    |
|----------------------------------------|----------------------------------|------------|-------------|----------------|-------------|------------------|---------------|----|
|                                        |                                  |            |             |                |             |                  | SEM           | OM |
| NBS [Stuckner <i>et al.</i> , 2022]    | Ni-based Superalloy              | Public     | Mask        | 512×512        | Unknown     | 23               | ✓             | ×  |
| UHCS-Seg [DeCost <i>et al.</i> , 2019] | Ultrahigh Carbon Steel           | Public     | Mask        | 645×475        | 1           | 24               | ✓             | ×  |
| DP590-1 [Li <i>et al.</i> , 2025]      | Dual-phase Steel                 | Private    | Mask        | 512×512        | 1           | 23               | ✓             | ×  |
| DP590-2 [Li <i>et al.</i> , 2025]      | Dual-phase Steel                 | Private    | Mask        | 1024×768       | 1           | 21               | ✓             | ×  |
| Carbide [Ma and Yu, 2023]              | Fe-0.2C-1.35Mn-2.5Cr-1.5Si Alloy | Private    | Mask        | 2048×1536      | 1           | 64               | ✓             | ×  |
| <b>IPSM-Bench (Ours)</b>               | <b>Zinc-based Alloys</b>         | This Paper | <b>Mask</b> | <b>512×512</b> | <b>20</b>   | <b>1,054</b>     | ✓             | ✓  |

Table 1: Comparison of IPSM-Bench with existing intermediate phase segmentation datasets.

named **IPSM-Bench**<sup>1</sup>, as shown in Figure 1. IPSM-Bench contains 1,054 microstructural images (512×512), comprising SEM and OM images, covering 20 types of zinc-based alloys and 22,179 manually annotated intermediate phase instances. All annotations are performed by experts in metallic materials science under quality control, ensuring precise delineation of intermediate phase boundaries. This dataset addresses a critical gap in the field by providing scarce, high-quality annotated data that serves as a reliable foundation for training and evaluating intermediate phase segmentation models. Furthermore, we propose a new Spatial Context Prior-guided SAM method, named **SCoP-SAM**, which extracts intermediate phase gradient structure and grayscale property priors as spatial context and integrates them into both the encoder and decoder of the SAM (Segment Anything Model) to provide prior-aware prompts for intermediate phase regions. Our SCoP-SAM precisely anchors the target intermediate phase regions during segmentation and fully delineates their complex boundary contours, thereby significantly improving segmentation accuracy.

The main contributions are summarized as follows: 1) To the best of our knowledge, we construct the first large-scale, diverse, and extensively annotated high-quality benchmark **IPSM-Bench** for intermediate phase segmentation in zinc alloy microstructures, providing a reliable foundation for train-

ing intermediate phase segmentation models and a fair comparison platform to drive methodological advances in analyzing zinc alloy microstructures. 2) We propose a new spatial context prior-guided SAM method (**SCoP-SAM**) that incorporates gradient-structure and grayscale-property priors into SAM to address key challenges in intermediate phase segmentation—low contrast, difficulty detecting small targets, and complex morphologies. 3) Extensive experiments are conducted on the proposed IPSM-Bench benchmark dataset and two other representative public alloy datasets to demonstrate the effectiveness of our SCoP-SAM method for intermediate phase segmentation in microstructure images.

## 2 Related Work

### 2.1 Metallic Microstructure Image Datasets

In the field of metallic microstructure image analysis, open-access, large, and finely annotated datasets for intermediate phases remain scarce. As shown in Table 1, NBS [Stuckner *et al.*, 2022] comprised only 23 intermediate phase particle images of nickel-based superalloys, with three segmentation classes: matrix phase, secondary precipitates, and tertiary precipitates. UHCS [DeCost *et al.*, 2017] focused on the pearlite-to-spheroidized cementite transformation in ultrahigh carbon steel but lacked pixel-level annotations; its segmented variant, UHCS-Seg [DeCost *et al.*, 2019], provided 24 images of intermediate phase particles. MetalDAM [Lu-

<sup>1</sup><https://github.com/AgileMotionTeam/IPSM-Bench>

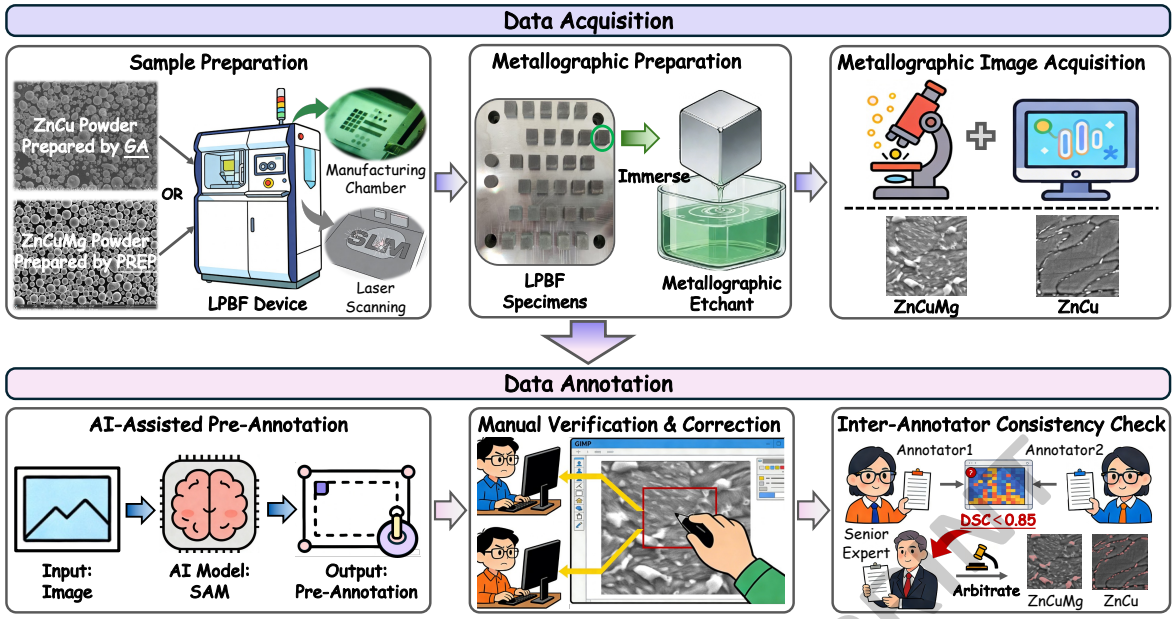


Figure 2: Data acquisition and annotation pipeline for IPSM-Bench.

engo *et al.*, 2022] contained 42 high-resolution SEM images of additively manufactured steel, with pixel-level annotations for five microstructural components; however, intermediate phases accounted for an extremely low proportion (about 0.24%), leading to incomplete annotations that hindered the training of intermediate phase segmentation models. Notably, recent work [Li *et al.*, 2025] has leveraged SAM for training-free microstructure analysis, but no standardized intermediate phase dataset has been established to date. Thus, there is an urgent need to develop a larger, more diverse, and more thoroughly annotated high-quality dataset for training intermediate phase segmentation models and as a fair comparison platform to advance research in this field.

## 2.2 SAM-based Image Segmentation Models in Multidisciplinary Applications

SAM [Kirillov *et al.*, 2023], a foundational segmentation model based on the prompt-learning paradigm, enables zero-shot segmentation and supports multiple prompt types, improving the accuracy of complex-structure analysis [Ma *et al.*, 2025; Abebe *et al.*, 2025; Guru *et al.*, 2025; Ma *et al.*, 2024; Marks *et al.*, 2025; Chen *et al.*, 2024; Wu *et al.*, 2025; Ji *et al.*, 2024; Wu and Liu, 2026]. Recently, SAM-Adapter [Chen *et al.*, 2023] developed a lightweight adaptation framework that converts task-specific knowledge into visual prompts to address poor generalization in specialized segmentation tasks. CAT-SAM [Xiao *et al.*, 2024] proposed a few-shot conditional tuning network via a prompt-bridging structure that enables decoder-guided collaborative tuning of the encoder to address SAM’s adaptation challenges. SAMCT [Lin *et al.*, 2025] was equipped with a U-shaped CNN image encoder based on SAM, which provides supplementary local features for segmentation. DenseSAM [Zhou *et al.*, 2025] replaced SAM’s reliance on positional prompts in dense scenarios with

semantic guidance, introducing an efficient semantic injection module and a dual-head decoding structure to tackle the challenge of dense objects in pathology and remote sensing images. MatSAM [Li *et al.*, 2025] developed a SAM-based automatic analysis model for material microstructures by integrating a prompt-generation strategy and multi-scale feature fusion, enabling efficient segmentation of structures such as grains and second-phase particles.  $\mu$ SAM [Archit *et al.*, 2025] provided a microscopy image segmentation and tracking tool based on SAM, unifying interactive and automatic segmentation across 2D/3D/temporal sequence data. Despite SAM’s efforts in microstructure analysis, existing methods still suffer from recognition accuracy for fine structures (e.g., grains, precipitates, microcracks) and a high susceptibility to misjudgment in complex phase-distribution scenarios.

## 3 Dataset: IPSM-Bench

### 3.1 Construction Necessity

Zinc-based absorbable biomaterials exhibit unique advantages and broad application prospects in medical fields, owing to a moderate degradation rate matching the repair cycle of human bone tissue, excellent biocompatibility, and favorable mechanical properties [Li *et al.*, 2020]. Therefore, we construct a new microstructure image dataset, named **IPSM-Bench**, spanning diverse zinc-based alloys to precisely analyze intermediate phases, which are critical for optimizing zinc alloy performance.

### 3.2 Data Acquisition

Our IPSM-Bench comprises 1,054 images, including 800 self-constructed and 254 sourced from the academic literature, covering two image types acquired via Optical Microscopy (OM) and Scanning Electron Microscopy (SEM). The data acquisition process is illustrated in Figure 2, which

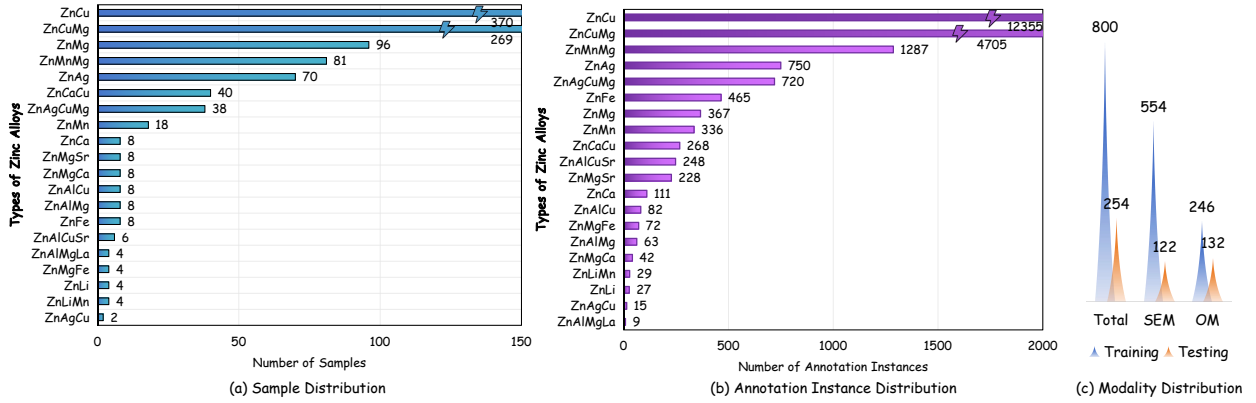


Figure 3: Statistical information of IPSM-Bench.

includes sample preparation, metallographic preparation, and metallographic image acquisition.

**Sample Preparation.** All samples were fabricated using the Laser Powder Bed Fusion (LPBF) technique. The raw materials were spherical zinc alloy powders produced by Gas Atomization (GA) and the Plasma Rotating Electrode Process (PREP). Fabrication was performed on a commercial LPBF system (SLM 125HL, SLM Solutions, Germany). Bulk solid samples were prepared by precisely adjusting laser processing parameters, with the laser energy density strictly maintained at 30-120 J/mm<sup>3</sup>. Ultimately, bulk samples measuring 10×10×10mm were obtained. After removing the surface oxide layer, the samples’ densities were measured using the Archimedes method. Only samples with densities above 99% were selected for subsequent metallographic preparation.

**Metallographic Preparation.** The surface parallel to the substrate was chosen for metallographic observation. Samples were subjected to a graded mechanical grinding process using 400#, 800#, 1500#, 3000#, and 5000# abrasive papers in sequence (#: grit of the sandpaper). After grinding, the surface was polished sequentially with W2.5μm, W1.5μm, and W0.5μm diamond metallographic polishing pastes until a bright, scratch-free surface suitable for microscopic observation was obtained. The polished observation surface was immersed in a dedicated enhanced metallographic etchant for 1-2 seconds, immediately removed, thoroughly rinsed with anhydrous ethanol, and dried prior to subsequent observation.

**Metallographic Image Acquisition.** OM images of the microstructures are acquired using an optical microscope (BX53m, Olympus, Japan), while SEM images are obtained with a cold-field emission scanning electron microscope (SU8100, Hitachi, Japan). In addition to the experimentally acquired images, literature-derived data were collected by downloading microstructure images reported in relevant academic literature. These images were further processed by removing interfering regions and retaining valid areas to ensure consistency and usability within the dataset.

### 3.3 Data Annotation

To ensure annotation efficiency, we adopt an AI-assisted annotation strategy in this work. The entire annotation process comprises three key stages: AI model-based pre-annotation,

manual verification and correction, and an inter-annotator consistency check.

**AI-Assisted Pre-Annotation.** A pre-trained segmentation model is used to generate initial segmentation masks for all images, thereby reducing the manual annotation workload. For zinc alloy intermediate phases with complex morphologies (e.g., lamellar, spherical, dendritic, and needle-like), SAM is selected as the AI auxiliary tool.

**Manual Verification and Correction.** We invite two professional annotators with more than 5 years of experience analyzing the microstructures of metallic materials to manually verify and correct the AI-generated pre-annotations. The correction criteria are based on the metallographic characteristics of zinc alloys: 1) *Boundary Correction*: For intermediate phases with blurred or inaccurate segmentation boundaries, the annotators adjusted the mask edges using professional annotation software (GIMP) to ensure precise delineation of intermediate phase contours. 2) *Missing and False Annotation Correction*: The annotators supplemented the missing small-scale intermediate phases that the AI model did not recognize and eliminated false annotations of matrix regions misidentified as intermediate phases.

**Inter-Annotator Consistency Check.** To ensure the reliability of the annotated data, an inter-annotator consistency check was conducted after manual correction. The Dice Similarity Coefficient (DSC) is used as an evaluation metric to quantify the consistency between the two annotators’ results. For annotations with DSC<0.85 (i.e., low consistency), a third senior materials science expert was invited to arbitrate and determine the final annotation. The average DSC of the final annotated dataset exceeds 0.90, indicating high annotation consistency and reliability.

### 3.4 Statistics

IPSM-Bench includes 20 Zinc-based alloy systems and comprises 1,054 images. Figure 3 (a) shows the distribution of Zinc alloy types, with ZnCu (370 samples, 35.10% of the total) and ZnCuMg (269 samples, 25.52% of the total) accounting for 60.63% of the total. IPSM-Bench has 22179 annotation instances. Figure 3 (b) illustrates the distribution of annotations across Zinc alloy types, with ZnCu (12355 annotation instances, 55.71% of the total) and ZnCuMg (4705 annota-

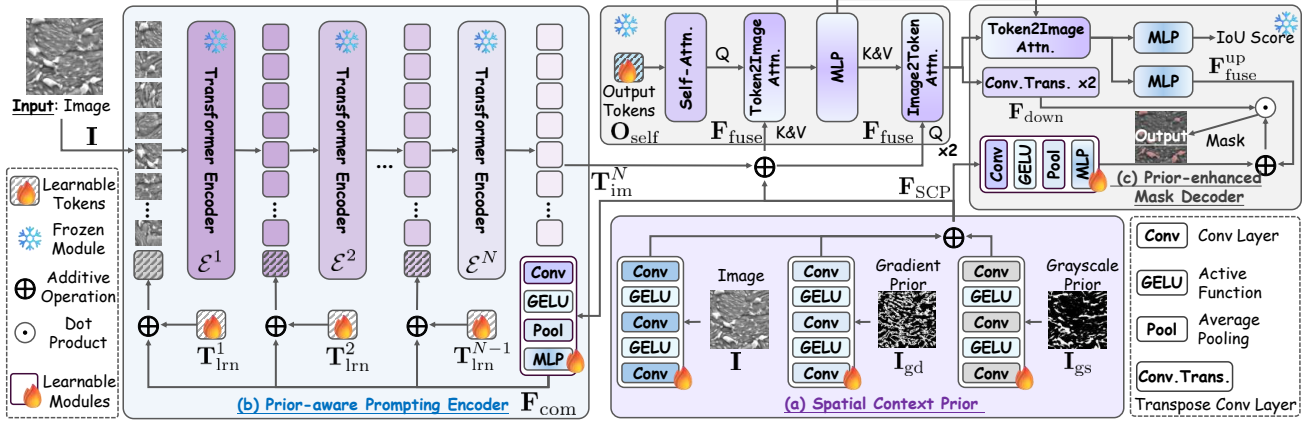


Figure 4: Architecture of the proposed SCoP-SAM. It comprises three core modules: (a) Spatial Context Prior, (b) Prior-aware Prompting Encoder, and (c) Prior-enhanced Mask Decoder.

tion instances, 21.21% of the total) accounting for 76.92% of the total. Figure 3 (c) presents the distribution of data modalities across training and testing. The training set includes 554 SEM images and 246 OM images, while the test set contains 122 SEM images and 132 OM images.

## 4 Methodology: SCoP-SAM

In this work, we propose a new Spatial Context Prior-guided SAM method (*SCoP-SAM*) for microstructure images. As shown in Figure 4, our SCoP-SAM comprises three core modules: Spatial Context Prior, Prior-aware Prompting Encoder, and Prior-enhanced Mask Decoder.

### 4.1 Spatial Context Prior (SCP)

To construct a spatial context prior for each input image, the SCP module integrates three types of information—RGB image, gradient prior, and grayscale prior—via three weight-sharing-free learnable networks. Each network comprises three convolutional layers interleaved with two GELU nonlinearities. Specifically, given an input image  $I$ , we introduce three learnable networks  $f_{im}$ ,  $f_{gd}$ , and  $f_{gs}$  to compute the spatial context prior  $F_{SCP}$  as follows:

$$F_{SCP} = f_{im}(I) + f_{gd}(I_{gd}) + f_{gs}(I_{gs}) \quad (1)$$

where  $F_{SCP}$  has a dimensionality of  $H_1 \times W_1 \times D_1$ . Here,  $I_{gd}$  (gradient prior) and  $I_{gs}$  (grayscale prior) are derived via edge detection and grayscale thresholding algorithms, respectively. For the gradient prior, we employ an adaptive edge-detection algorithm to generate a gradient structure map, integrating the Canny operator and the Otsu thresholding method to extract edge features. Then, by comparing contour counts, the edge map with a greater number of contours is selected as  $I_{gd}$ , aiming to maximize the preservation of topological integrity and structural continuity of the intermediate phase edges. For the grayscale prior, we model the grayscale distribution using dominant color analysis, yielding a grayscale map  $I_{gs}$ . Specifically, we first analyze the distribution of black and white pixels in  $I_{im}$  and determine the target dominant color (black or white) based on the dominant pixel count. Subsequently, K-means clustering is applied to all image pixels, and

the cluster center closest to the target dominant color is selected via Euclidean distance. Finally, a binary regional mask is generated around the cluster center, effectively capturing the aggregation properties of the intermediate phase region in the grayscale space.

### 4.2 Prior-aware Prompting Encoder (PPE)

Upon obtaining  $F_{SCP}$ , we embed it into the subsequent Prior-aware Prompting Encoder module to generate a robust prior prompt that guides downstream image encoding. Specifically, we introduce a learnable projection network  $f_{proj}$  to project  $F_{SCP}$  onto common prompting tokens  $F_{com}$ . Subsequently, we add  $F_{com}$  with  $L_p$  learnable tokens  $T_{lrn}^i$  to generate prior-aware prompting tokens  $T_{prio}^i$  for feeding into the  $i$ -th transformer encoder  $\mathcal{E}^i$ . This process can be formulated as:

$$F_{com} = f_{proj}(F_{SCP}), \quad (2)$$

$$T_{prio}^i = F_{com} \oplus T_{lrn}^i, \quad i = 1, \dots, N-1 \quad (3)$$

$$T_{all}^i = \mathcal{E}^i(\text{Concat}(T_{im}^i, T_{prio}^i)), \quad i = 1, \dots, N-1, \quad (4)$$

$$T_{im}^N = \mathcal{E}^N(T_{all}^{N-1}), \quad (5)$$

where  $f_{proj}$  is a learnable network consisting of a convolutional layer, GELU, average pooling, and an MLP. Notably, the PPE module incorporates  $N$  transformer encoders  $\{\mathcal{E}^i\}_{i=1}^N$ , where  $\mathcal{E}^{i+1}$  encodes the output of  $\mathcal{E}^i$ , updating the prior-aware prompting token input from  $T_{prio}^i$  to  $T_{prio}^{i+1}$  during encoding. For the final transformer encoder  $\mathcal{E}^N$ , we retain only  $T_{im}^N$  (i.e., the prior-enhanced image embedding tokens) and exclude  $T_{prio}^N$  from its output.

### 4.3 Prior-enhanced Mask Decoder (PMD)

The PMD module integrates a spatial context prior and maps the prior-aware, token-guided image embedding to an output mask. Specifically, after obtaining the image embedding  $T_{im}^N$ , we add the spatial context prior  $F_{SCP}$  to it and feed their integration  $F_{fuse}$  into a two-block decoder  $\mathcal{D}_1$ . Each block of  $\mathcal{D}_1$  contains a self-attention layer, a cross-attention (token-to-image) layer, an MLP, and a cross-attention (image-to-token)

| Method                                  | Venue              | Backbone  | IPSM-Bench (%) |              |              |
|-----------------------------------------|--------------------|-----------|----------------|--------------|--------------|
|                                         |                    |           | mIoU           | mF1          | mPre         |
| SAM [Kirillov <i>et al.</i> , 2023]     | ICCV'23            | SAM ViT-H | 29.05          | 44.83        | 60.98        |
| SAM-Adapter [Chen <i>et al.</i> , 2023] | ICCVW'23           | SAM ViT-H | 55.12          | 63.50        | 70.49        |
| SAMCT [Lin <i>et al.</i> , 2025]        | ICLR'24            | SAM ViT-H | <u>62.01</u>   | 69.14        | 71.13        |
| CAT-SAM [Xiao <i>et al.</i> , 2024]     | ECCV'24            | SAM ViT-H | 43.05          | 55.60        | 34.87        |
| DenseSAM [Zhou <i>et al.</i> , 2025]    | IJCAI'25           | SAM ViT-H | 59.82          | <u>73.13</u> | <u>84.79</u> |
| MatSAM [Li <i>et al.</i> , 2025]        | Acta Materialia'25 | SAM ViT-H | 29.01          | 44.78        | 61.02        |
| $\mu$ SAM [Archit <i>et al.</i> , 2025] | Nature Methods'25  | SAM ViT-H | 39.19          | 44.08        | 42.48        |
| <b>SCoP-SAM (Ours)</b>                  | This Paper         | SAM ViT-H | <b>71.52</b>   | <b>82.82</b> | <b>87.55</b> |

Table 2: Comparison results against state-of-the-art methods on IPSM-Bench. The best and second-best results are in **bold** and underlined.

| Method                                  | Venue              | NBS (%)      |              |              | UHCS (%)     |              |              |
|-----------------------------------------|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                                         |                    | mIoU         | mF1          | mPre         | mIoU         | mF1          | mPre         |
| SAM [Kirillov <i>et al.</i> , 2023]     | ICCV'23            | 62.54        | 76.86        | 79.11        | 32.00        | 45.66        | 52.60        |
| SAM-Adapter [Chen <i>et al.</i> , 2023] | ICCVW'23           | 26.82        | 36.78        | 63.24        | <u>44.40</u> | <u>48.46</u> | 66.75        |
| SAMCT [Lin <i>et al.</i> , 2025]        | ICLR'24            | 34.50        | 43.80        | 49.19        | 43.55        | 46.62        | <b>90.50</b> |
| CAT-SAM [Xiao <i>et al.</i> , 2024]     | ECCV'24            | 24.73        | 33.50        | 55.12        | 43.48        | 46.51        | 46.01        |
| DenseSAM [Zhou <i>et al.</i> , 2025]    | IJCAI'25           | 29.18        | 40.78        | 74.18        | 37.59        | 48.36        | 49.94        |
| MatSAM [Li <i>et al.</i> , 2025]        | Acta Materialia'25 | <u>62.73</u> | <u>77.00</u> | <u>79.36</u> | 33.04        | 47.15        | 54.34        |
| $\mu$ SAM [Archit <i>et al.</i> , 2025] | Nature Methods'25  | 31.10        | 41.38        | 46.76        | 43.48        | 46.51        | 43.48        |
| <b>SCoP-SAM (Ours)</b>                  | This Paper         | <b>67.26</b> | <b>80.42</b> | <b>80.43</b> | <b>57.25</b> | <b>67.48</b> | <u>80.93</u> |

Table 3: Comparison results against state-of-the-art methods on NBS and UHCS. The best and second-best results are in **bold** and underlined.

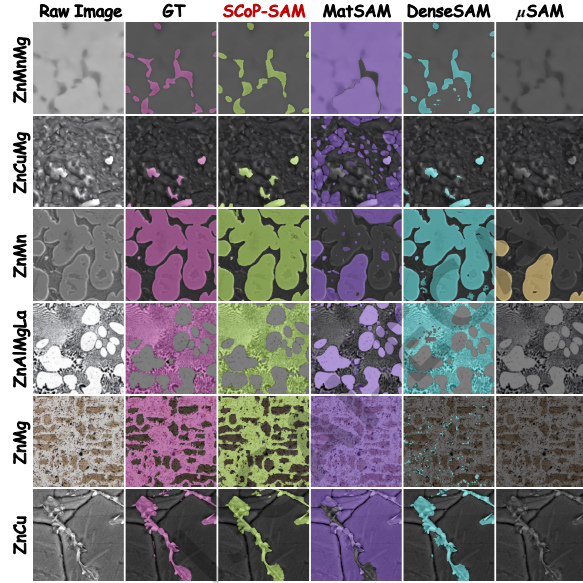


Figure 5: Visualization of the proposed SCoP-SAM.

layer. A set of learnable output tokens  $\mathbf{O}_{\text{self}}$  undergoes self-attention and serves as a query to attend to the prior-enhanced image embedding  $\mathbf{F}_{\text{fuse}}$ , enabling it to capture global image context.  $\mathbf{F}_{\text{fuse}}$  acts as the key and value for the token-to-image cross-attention, and the refined output token  $\mathbf{O}_{\text{self}}^{\text{re}}$  is updated via an MLP to integrate image contextual information.  $\mathbf{F}_{\text{fuse}}$  serves as a query to attend to the updated output token via the image-to-token cross-attention, injecting global guidance

into local spatial features and obtaining the prior-enhanced image embedding  $\mathbf{F}_{\text{fuse}}^{\text{en}}$ . After a two-block decoder  $\mathcal{D}_1$ ,  $\mathbf{F}_{\text{fuse}}^{\text{en}}$  is upsampled via two transposed convolutions with GELU activations and layer normalization, yielding a downsampled feature map  $\mathbf{F}_{\text{down}}$  relative to the input image. The refined output token  $\mathbf{O}_{\text{self}}^{\text{re}}$  is fed into a token-to-image cross-attention layer, followed by an MLP that produces an upsampled image embedding  $\mathbf{F}_{\text{fuse}}^{\text{up}}$ . The point-wise sum of  $\mathbf{F}_{\text{fuse}}^{\text{up}}$  and  $\mathbf{F}_{\text{com}}$  (i.e.,  $f_{\text{com}}(\mathbf{F}_{\text{SCP}})$ ) followed by the point-wise multiplication of  $\mathbf{F}_{\text{down}}$  and  $\mathbf{F}_{\text{com}}$  yields the foreground probability for each pixel, producing the final segmentation mask.

#### 4.4 Training

The overall training objective of SCoP-SAM is formulated as a linearly weighted multi-task loss function that integrates three functionally complementary core loss components: a region-overlap optimization component (Dice loss), a pixel-wise binary classification component (binary cross-entropy loss), and an auxiliary intersection-over-union (IoU) regression component for segmentation consistency.

## 5 Experiments

### 5.1 Experimental Setup

**Datasets.** To evaluate the proposed SCoP-SAM, we conduct experiments on various datasets, including IPSM-Bench, NBS [Stuckner *et al.*, 2022], and UHCS [Na *et al.*, 2023]. All methods are trained on the IPSM-Bench training set and evaluated on the test sets of IPSM-Bench, NBS, and UHCS. This design is due to the training sets for NBS and UHCS being too small to support model training. Specifically, for IPSM-Bench, the training set comprises 800 samples: 246 from the

OM modality and 554 from the SEM modality, while the test set comprises 254 samples: 132 from the OM modality and 122 from the SEM modality. More details are provided in Figure 3 and subsection 3.4. UHCS has 5 labeled SEM images of cementite particles for testing. NBS has 9 labeled SEM images of nisuperalloy for testing.

**Metrics.** Following prior efforts [Zhou *et al.*, 2025; Li *et al.*, 2025; Chen *et al.*, 2023], we adopt three pixel-level metrics: Intersection over Union (IoU), F1 score, and Precision. Specifically, we compute the performance for the foreground and background classes separately for each metric and report the mean across two classes, denoted mIoU, mF1, and mPre.

**Implementation Details.** To ensure the fairness of comparative experiments, all experiments adhere to consistent configurations: 1) all methods use SAM ViT-H [Kirillov *et al.*, 2023] as the backbone; 2) all methods use their official open-source implementations and settings, trained from scratch on IPSM-Bench’s training set. We train SCoP-SAM on four NVIDIA A800 GPUs with a batch size of 4 for 50 epochs using the AdamW optimizer and a learning rate of  $1e-4$ .

## 5.2 Results and Analysis

**Performance Comparisons on IPSM-Bench.** We compare our SCoP-SAM with SOTA methods [Zhou *et al.*, 2025; Xiao *et al.*, 2024; Chen *et al.*, 2023; Lin *et al.*, 2025; Archit *et al.*, 2025; Kirillov *et al.*, 2023; Li *et al.*, 2025] on IPSM-Bench. In Table 2, SCoP-SAM consistently outperforms all methods. For instance, SCoP-SAM improves mIoU by 11.70% over DenseSAM, 9.51% over SAMCT, and 28.47% over CAT-SAM. We attribute these improvements primarily to incorporating a spatial context prior and prior-aware prompts into the learnable modules of both the encoder and decoder, enabling robust performance under the challenges of intermediate phase segmentation, including low inter-phase contrast, blurred or incomplete boundaries, a high proportion of small-scale intermediate phases, heterogeneous morphologies, and interference from noise.

**Performance Comparisons on NBS and UHCS.** We further investigate the generalization of SCoP-SAM in few-shot settings with pronounced distribution shifts in materials data, such as large variations in precipitate size and more drastic morphological changes. To this end, we evaluate our SCoP-SAM on the more challenging NBS dataset and report the results in Table 3. Notably, the original SAM already achieves 62.54% on NBS, whereas several SAM-based variants (i.e., SAM-Adapter, SAMCT, CAT-SAM, DenseSAM, and  $\mu$ SAM) exhibit performance degradation, suggesting that their generalization under severe domain shifts remains limited. In contrast, SCoP-SAM outperforms MatSAM by 4.53%, further demonstrating the robustness of our method beyond the training data distribution. We attribute this advantage primarily to SCoP-SAM’s use of multi-source spatial context priors to better exploit structural cues, enabling more stable object localization and boundary delineation under large-scale variations and morphological heterogeneity, thereby mitigating performance drops caused by few-shot learning and cross-domain distribution discrepancies. For the UHCS dataset, SCoP-SAM outperforms DenseSAM, SAM-Adapter, and SAMCT by 19.66%, 12.85%, and 13.70% in

| Module                      | mIoU (%)     | mF1 (%)      | mPre (%)     |
|-----------------------------|--------------|--------------|--------------|
| Baseline (ViT-H)            | 29.05        | 44.83        | 60.98        |
| Baseline+SCP                | 48.18        | 61.72        | 69.98        |
| Baseline+SCP+PPE            | 59.49        | 72.98        | 81.38        |
| <b>Baseline+SCP+PPE+PMD</b> | <b>71.52</b> | <b>82.82</b> | <b>87.55</b> |

Table 4: Ablation study results on IPSM-Bench. Our strategy is highlighted in **bold**.

mIoU, respectively, and demonstrates strong generalization, indicating that our IPSM-Bench also covers diverse intermediate phase morphologies and imaging conditions with high representativeness. Notably, SAMCT achieves a higher mPre (90.50%) than SCoP-SAM (80.93%) on UHCS; yet it fails to achieve a higher mIoU. This is because SAMCT has a higher miss rate (i.e., false negatives), predicting only a small set of high-confidence regions, which inflates precision, but insufficient mask coverage substantially degrades IoU.

## 5.3 Ablation Study

We evaluate the effects of three core modules (i.e., SCP, PPE, and PMD). In Table 4, SCP alone achieves 48.18% mIoU; adding PPE improves performance to 59.49%; integrating PMD into the complete model yields optimal performance of 71.52%. These results demonstrate that each module positively contributes to segmentation performance and that the modules work synergistically to enhance overall performance, supporting SCoP-SAM’s performance advantage.

## 5.4 Visualization Analysis

We provide a visual analysis of SEM and OM images of six typical zinc alloys, as shown in Figure 5. The SEM and OM images present distinct challenges: SEM images suffer from low contrast and background noise, while OM images exhibit complex topological structures of microstructural phases. Compared with the problems of over-segmentation, under-segmentation, or even complete failure in MatSAM, DenseSAM, and  $\mu$ SAM, our SCoP-SAM can stably and accurately restore irregular morphology and detailed boundaries, achieving high consistency with ground-truth (GT) annotations across all samples and demonstrating excellent cross-material and cross-modality generalization and robustness.

## 6 Conclusion

In this paper, we present IPSM-Bench, a new, large-scale, diverse, and high-quality benchmark for intermediate phase segmentation in zinc alloy microstructures, providing a reliable foundation for training and testing segmentation models. We further propose SCoP-SAM, which incorporates spatial context priors into image encoding and mask decoding generation, enabling more reliable use of the gradient structure and grayscale properties while improving both segmentation performance and boundary quality. Extensive experiments demonstrate that SCoP-SAM achieves state-of-the-art results on IPSM-Bench and delivers clear gains on two other public alloy benchmarks, indicating strong segmentation ability across various alloy scenarios.

## Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (52541014, 52541013, 62522102, 52522110, 62432001), Beijing Natural Science Foundation (4252020, L247006), Fundamental Research Funds for the Central Universities (FRF-EF-25-05A), and Young Elite Scientists Sponsorship Program of the Beijing High Innovation Plan (20250747).

## References

- [Abebe *et al.*, 2025] Waqwoya Abebe, Jan Strube, Luanzheng Guo, Nathan R Tallent, Oceane Bel, Steven Spurgeon, Christina Doty, and Ali Jannesari. Sam-i-am: Semantic boosting for zero-shot atomic-scale electron micrograph segmentation. *Computational Materials Science*, 246:113400, 2025.
- [Archit *et al.*, 2025] Anwai Archit, Luca Freckmann, Sushmita Nair, Nabeel Khalid, Paul Hilt, Vikas Rajashekar, Marei Freitag, Carolin Teuber, Melanie Spitzner, Constanza Tapia Contreras, et al. Segment anything for microscopy. *Nature Methods*, 22(3):579–591, 2025.
- [Chen *et al.*, 2023] Tianrun Chen, Lanyun Zhu, Chaotao Deng, Runlong Cao, Yan Wang, Shangzhan Zhang, Zejian Li, Lingyun Sun, Ying Zang, and Papa Mao. Sam-adapter: Adapting segment anything in underperformed scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3367–3375, 2023.
- [Chen *et al.*, 2024] Wei-Ting Chen, Yu-Jiet Vong, Sy-Yen Kuo, Sizhou Ma, and Jian Wang. Robustsam: Segment anything robustly on degraded images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4081–4091, 2024.
- [DeCost *et al.*, 2017] Brian L DeCost, Matthew D Hecht, Toby Francis, Bryan A Webler, Yoosuf N Picard, and Elizabeth A Holm. Uhcsdb: ultrahigh carbon steel micrograph database: tools for exploring large heterogeneous microstructure datasets. *Integrating Materials and Manufacturing Innovation*, 6(2):197–205, 2017.
- [DeCost *et al.*, 2019] Brian L DeCost, Bo Lei, Toby Francis, and Elizabeth A Holm. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel. *Microscopy and Microanalysis*, 25(1):21–29, 2019.
- [Guru *et al.*, 2025] Mahish K Guru, Jan Bohlen, Roland C Aydin, and Noomane Ben Khalifa. Machine learning pipeline for structure-property modeling in mg-alloys using microstructure and texture descriptors. *Acta Materialia*, page 121132, 2025.
- [Huang *et al.*, 2025] Chengcong Huang, Yizhu Wang, Fan Yang, Yixuan Shi, Shangyan Zhao, Xuan Li, Yuchen Lu, Yuzhi Wu, Jie Zhou, Amir A Zadpoor, et al. Additively manufactured biodegradable zn-mn-based implants with an unprecedented balance of strength and ductility. *Acta Biomaterialia*, 196:506–522, 2025.
- [Ji *et al.*, 2024] Wei Ji, Jingjing Li, Qi Bi, Tingwei Liu, Wenbo Li, and Li Cheng. Segment anything is not always perfect: An investigation of sam on different real-world applications. *Machine Intelligence Research*, 21(4):617–630, 2024.
- [Kirillov *et al.*, 2023] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.
- [Li *et al.*, 2020] Yageng Li, Holger Jahr, Jie Zhou, and Amir Abbas Zadpoor. Additively manufactured biodegradable porous metals. *Acta Biomaterialia*, 115:29–50, 2020.
- [Li *et al.*, 2025] Changtai Li, Xu Han, Chao Yao, Yu Guo, Zixin Li, Lei Jiang, Wei Liu, Haiyou Huang, Huadong Fu, and Xiaojuan Ban. A novel training-free approach to efficiently extracting material microstructures via visual large model. *Acta Materialia*, 290:120962, 2025.
- [Lin *et al.*, 2025] Xian Lin, Yangyang Xiang, Zhehao Wang, Kwang-Ting Cheng, Zengqiang Yan, and Li Yu. Samct: Segment anything allowing labor-free task-indicator prompts. *IEEE Transactions on Medical Imaging*, 44(3):1386–1399, 2025.
- [Luengo *et al.*, 2022] Julian Luengo, Raul Moreno, Ivan Sevillano, David Charte, Adrian Pelaez-Vegas, Marta Fernandez-Moreno, Pablo Mesejo, and Francisco Herrera. A tutorial on the segmentation of metallographic images: Taxonomy, new metaldam dataset, deep learning-based ensemble model, experimental analysis and challenges. *Information Fusion*, 78:232–253, 2022.
- [Ma and Yu, 2023] Xudong Ma and Yunhe Yu. Training tricks for steel microstructure segmentation with deep learning. *Processes*, 11(12):3298, 2023.
- [Ma *et al.*, 2024] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- [Ma *et al.*, 2025] Xudong Ma, Yuqi Zhang, Chenchong Wang, and Wei Xu. Alloy microstructure segmentation through sam and domain knowledge without extra training. *Scripta Materialia*, 260:116581, 2025.
- [Marks *et al.*, 2025] Markus Marks, Uriah Israel, Rohit Dilip, Qilin Li, Changhua Yu, Emily Laubscher, Ahamed Iqbal, Elora Pradhan, Ada Ates, Martin Abt, et al. Cell-sam: a foundation model for cell segmentation. *Nature Methods*, pages 1–9, 2025.
- [Na *et al.*, 2023] Juwon Na, Se-Jong Kim, Heekyu Kim, Seong-Hoon Kang, and Seungchul Lee. A unified microstructure segmentation approach via human-in-the-loop machine learning. *Acta Materialia*, 255:119086, 2023.
- [Stuckner *et al.*, 2022] Joshua Stuckner, Bryan Harder, and Timothy M Smith. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *npj Computational Materials*, 8(1):200, 2022.

- [Tong *et al.*, 2023] Xian Tong, Yue Han, Li Zhu, Runqi Zhou, Zhiqiang Lin, Hongning Wang, Shengbin Huang, Yuncang Li, Jianfeng Ma, Cuie Wen, et al. Znp-coated zn-1cu-0.1ti membrane with high strength-ductility, antibacterial ability, cytocompatibility, and osteogenesis for biodegradable guided bone regeneration applications. *Advanced Functional Materials*, 33(31):2214657, 2023.
- [Wu and Liu, 2026] Xin-Jian Wu and Cheng-Lin Liu. Hierarchical open-vocabulary part-object segmentation with knowledge-guided sam. *Machine Intelligence Research*, 23(1):214–226, 2026.
- [Wu *et al.*, 2025] Jinxuan Wu, Jiange Wang, and Dongdong Zhang. Gausam: Contour-guided 2d gaussian fields for multi-scale medical image segmentation with segment anything. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
- [Xiao *et al.*, 2024] Aoran Xiao, Weihao Xuan, Heli Qi, Yun Xing, Ruijie Ren, Xiaoqin Zhang, Ling Shao, and Shijian Lu. Cat-sam: Conditional tuning for few-shot adaptation of segment anything model. In *European Conference on Computer Vision*, pages 189–206, 2024.
- [Yang *et al.*, 2023] Hongtao Yang, Dawei Jin, Jiansun Rao, Jiahui Shi, Guannan Li, Cheng Wang, Kai Yan, Jing Bai, Guo Bao, Meng Yin, et al. Lithium-induced optimization mechanism for an ultrathin-strut biodegradable zn-based vascular scaffold. *Advanced Materials*, 35(19):2301074, 2023.
- [Zhou *et al.*, 2025] Linyun Zhou, Jiacong Hu, Shengxuming Zhang, Xiangtong Du, Mingli Song, Xiuming Zhang, and Zunlei Feng. Densesam: Semantic enhance sam for efficient dense object segmentation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 7994–8002, 2025.