*metals*

MDPI

*Article*

# MaterialAlphaSAM: An Adaptive Prompting and Domain Adaptation-Based Segmentation Method for the Microstructure of Complex Titanium Alloys

**Ke Li [1], Bowen Deng [1], Yanru Zhao [1], Wei Liu [1,*], Chao Yang [1], Jing Zhu [1], Di Tie [2,*], Huixian Gao [1] and Wenzhong Luo [1]**

[1] Western Superconducting Technologies Co., Ltd., Xi'an 710018, China; like@c-wst.com (K.L.)
[2] School of Materials Science and Engineering, Guangdong Ocean University, Zhanjiang 529500, China
* Correspondence: liuwei421@126.com (W.L.); tie-di@hotmail.com (D.T.)

**Abstract**

Precise segmentation of high-magnification titanium alloy micrographs under few-shot scenarios remains a non-trivial task, primarily owing to the intricate morphology, heterogeneous discrete distribution, and weak phase boundaries of the primary α phase. To address these issues, this paper presents MaterialAlphaSAM, a lightweight domain-adaptive segmentation framework built upon the Segment Anything Model (SAM). Leveraging SAM's powerful global context modeling capability, the proposed method incorporates two key modules: a Geometry-Constrained Prompt Prior (GCPP) module and a Domain-Adaptation Adapter (DAA) module. The GCPP module explicitly embeds geometric and morphological priors to generate semantically guided prompts, effectively alleviating prompt redundancy and noise sensitivity. The DAA module performs cross-domain alignment of the encoder features, reducing the domain discrepancy between natural images and metallic microstructures. Extensive experiments demonstrate that both modules consistently boost segmentation performance. On the titanium alloy dataset, MaterialAlphaSAM achieves 89.53% IoU and a 94.40% F1-score, outperforming FCN, UNet, DeepLabV3, PSPNet and the vanilla SAM. It exhibits superior robustness to weak boundaries, fine-scale α phases, and complex background interference.

**Keywords:** segment anything model; titanium alloy; microstructure; segmentation method

## 1. Introduction

In the field of aeronautical manufacturing, titanium alloys have become the core material for key load-bearing components of aircraft due to their high specific strength, excellent resistance to extreme temperatures, and superior corrosion resistance. To ensure the structural integrity and reliability of titanium alloy components in complex service environments, their material properties must comply with stringent aviation quality standards. In accordance with the requirements of SAE AS9100, an international aerospace quality management system, titanium alloy parts are subject to multidimensional and full-process quality control. Among these methods, microstructural characterization, as a critical method for evaluating the metallurgical quality and mechanical properties of materials, runs through all stages, including raw material acceptance, process validation and finished product inspection. In particular, for duplex titanium alloys such as TC4 and TC11, their room-temperature strength, fracture toughness, fatigue crack growth rate and

Check for updates

Academic Editor: Chris Aldrich

Received: 15 May 2026
Revised: 14 June 2026
Accepted: 23 June 2026
Published: 2 July 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

high-temperature stability are highly dependent on their microstructural characteristics. As such, systematic metallographic analysis must be conducted to achieve precise regulation of microstructural parameters.

In practical applications, quantitative analysis of the primary α phase in titanium alloys is primarily based on metallographic microscopic image processing, which focuses on accurate segmentation and statistical characterization of the primary α phase within microstructures. According to the underlying methodology, relevant techniques fall into two main technical approaches: traditional image processing-based metallographic analysis and deep learning-based intelligent segmentation.

Traditional metallographic image processing methods typically follow a three-stage analytical workflow consisting of image preprocessing, phase region segmentation and quantitative analysis. As a fundamental approach for the determination of primary α phase content, such methods still hold significant practical value for samples with distinct microstructural characteristics [1–4]. First, the quality of raw images is improved to lay a foundation for subsequent segmentation via measures including median or Gaussian filtering for noise suppression, histogram equalization for contrast enhancement and homomorphic filtering for illumination non-uniformity correction. Subsequently, phase segmentation is achieved by threshold segmentation (e.g., Otsu's algorithm), edge detection (e.g., Canny operator), region growing or morphological processing, based on differences in gray level, edge, or regional characteristics between the primary α phase and the β matrix. Finally, representative volume fraction data are obtained by calculating the pixel proportion of the α phase in binary images, conducting result screening in combination with morphological parameters such as circularity and performing multi-field average processing. This measurement approach offers advantages such as clear underlying principles, high computational efficiency, strong interpretability and compatibility with the current metallographic standard system, and is therefore widely applied in routine quality inspection and industrial practice.

However, in the context of complex microstructures commonly encountered in aerospace manufacturing, such as blurred α phase boundaries in duplex microstructures, lath-shaped α phase adhesion after deformation and a high fraction of secondary α phase, the limitations of traditional methods have become increasingly prominent. The core limitations primarily manifest in the following four aspects. First, there is a strong reliance on image preprocessing. Metallographic images are often affected by uneven illumination, noise interference and blurred boundaries, and conventional methods such as threshold segmentation and edge detection require elaborate preprocessing procedures. Furthermore, the selection of preprocessing parameters is highly dependent on manual experience, and minor adjustments to parameters may lead to deviations in the analytical results. Second, the accuracy of microstructure boundary segmentation is inadequate. When metallographic microstructures involve multiphase interlocking and intercalation or insignificant grayscale differences at phase interfaces, traditional methods, including threshold segmentation and region growing, struggle to accurately distinguish the boundaries between different phases, which easily causes over-segmentation and under-segmentation, and directly compromising the accuracy of quantitative parameters. Third, the robustness to interference is weak. For metallographic images containing impurities, scratches and voids, traditional methods fail to distinguish target microstructures effectively from interfering substances and tend to misclassify interference regions as microstructural phases, resulting in distorted quantitative results. Fourth, the generality and degree of automation are relatively low. Metallographic microstructures vary drastically with different material grades and heat treatment processes, and traditional methods require the design of dedicated algorithmic procedures for specific microstructure types, making it difficult to adapt to diverse ana-

lytical requirements. Moreover, the whole analysis process involves numerous manual intervention steps, which precludes the realization of large-scale batch automatic analysis. Traditional image processing methods are applicable to titanium alloy samples with simple microstructures and well-defined contrasts. Nevertheless, against the backdrop of the growing demand for the segmentation of complex microstructural morphologies with high precision and high consistency, their robustness and adaptability have evidently become insufficient.

In recent years, with the rapid development of computer technology, deep learning-based methods have been widely applied to image segmentation of metallic material microstructures, enabling automated and quantitative microstructure characterization. Zhao et al. [4] established a semantic segmentation model for titanium alloy microstructures via U-Net, which successfully achieved multiphase segmentation of microstructure images for 15 types of titanium alloys. Biswas et al. [5] proposed a U-Net ensemble model integrated with an attention mechanism to segment metallic microstructure images under various color space transformations, including RGB, HSV and YUV, achieving a relatively high intersection over union of approximately 0.68 on the public MetalDAM dataset. Majumdar et al. [6] further improved U-Net to develop the MIMU-Net, in which recurrent residual blocks were incorporated into the encoder, and the spatial-channel squeeze-and-excitation module was fused with feature pyramid pooling to enhance the multi-scale segmentation capability. This model exhibited excellent performance on the MetalDAM dataset, outperforming a variety of existing methods. Baskaran et al. [7] proposed a two-stage approach combining convolutional neural network (CNN) classification with traditional image processing for the quantitative analysis of Ti-6Al-4V alloy microstructures. First, a three-convolutional-layer CNN was employed to classify the microstructures into three categories, (lamellar, duplex, and acicular) with an accuracy of 93%. Subsequently, specific algorithms were selected according to the classification results for phase content measurement, where the measurement error of primary $\alpha$-phase content was less than 2%. Moreover, several studies have combined deep semantic segmentation with traditional algorithms. The boundary class semantic segmentation (BCSS) method proposed by Fotos et al. [8] introduced boundary categories on the basis of U-Net segmentation and integrated the marker-based watershed algorithm to achieve accurate identification and segmentation of $\alpha$ phase, $\beta$ phase and grain boundaries in Ti-6Al-4V duplex microstructures. Comparative analysis with manual annotation in the original study demonstrated that the error in the $\alpha$ phase volume fraction obtained via this method was generally less than 10%, as specified by the ASTM standard, indicating the reliability of the measurement of the primary $\alpha$ phase volume fraction. In addition, to address the scarcity of annotated data, recent studies have explored the combination of deep learning method and generative models. Thool et al. [9] adopted U-Net and LinkNet with a ResNet-34 backbone for the segmentation of $\alpha/\beta$ and other phases in the complex multiphase microstructures of zinc alloys, assisted by preprocessing steps of anisotropic diffusion and contrast stretching, and attained an IoU score of over 85% for all phases. Gerçek et al. [10] proposed a Self-labeling Distillation U-Net, which utilized the Segment Anything Model to generate pseudo-labels and distill them into U-Net, thus achieving high-precision microstructure segmentation with only a small number of annotated images.

In summary, deep learning models based on the encoder–decoder architecture and its variants have achieved remarkable progress in the microstructure segmentation of titanium alloys and other multiphase metallic materials. Such methods have demonstrated outstanding performance in terms of segmentation accuracy and stability under complex microstructural conditions. Nevertheless, the aforementioned approaches generally rely on supervised learning frameworks tailored to specific material systems, etching conditions

and imaging modalities, which results in significant limitations in their application across different alloy compositions, processing states and imaging modalities. Furthermore, the heavy reliance of these methods on large-scale, high-quality pixel-level annotated data further constrains their practical application scope and efficiency.

In recent years, general visual foundation models represented by the Segment Anything Model (SAM) [11] have attracted extensive attention in the field of computer vision and introduced a new paradigm for material microstructure image analysis. Through pre-training on large-scale, multi-source datasets (e.g., the SA-1B dataset, which contains approximately 1.1 billion high-quality mask annotations) such models acquire visual representations with strong transferability and generalization, thereby exhibiting substantial segmentation potential under zero-shot or few-shot conditions. Theoretically, these foundation models possess universal segmentation capability across various tasks and object categories, offering a new feasible approach to reducing the reliance of specialized research fields on large-scale manually annotated data.

Nevertheless, the segmentation of the primary $\alpha$ phase in high-magnification microstructure images exhibits remarkable domain specificity in terms of imaging characteristics, morphological features and semantic definitions. On the one hand, the primary $\alpha$ phase in high-magnification microstructures presents diverse morphologies such as equiaxed, short rod-like and irregular polygonal shapes, with a highly discrete size distribution. The equivalent diameters typically show a prominent multimodal characteristic, featuring the coexistence of fine particles and coarse equiaxed grains and a size span of more than one order of magnitude. On the other hand, owing to the constraints of etching conditions and imaging mechanisms, the phase boundaries between the primary $\alpha$ phase and transformed $\beta$ matrix often present the characteristics of low contrast and gradual transition, with boundary sharpness significantly lower than that of common object contours in natural images, thus increasing the difficulty of accurate segmentation.

Furthermore, existing studies have demonstrated [12,13] that the SAM exhibits high sensitivity to the spatial distribution of input prompts and the implicit semantic priors they contain, and its segmentation results largely depend on whether the prompts can effectively indicate the target regions. In microstructure images, owing to the highly discrete sizes, inhomogeneous spatial distributions and distinct low-contrast boundaries of the primary $\alpha$ phase, grid prompts with a fixed density struggle to capture fine particles and complex boundary regions simultaneously under few-shot conditions. Sparse sampling tends to overlook the fine-scale $\alpha$ phase; on the other hand, blind elevation of the sampling density will lead to many invalid or semantically ambiguous prompts, compromising the target region discrimination capability of the model. This issue is particularly pronounced in lightweight adaptation scenarios, emerging as a key bottleneck that restricts the performance of general foundation models in microstructure segmentation. Against this backdrop, how to incorporate the domain priors of microstructures to adaptively generate high-quality prompts with spatial and semantic consistency has gradually become a core issue restricting the application performance of general foundation models in this field. Accordingly, how to adaptively generate high-quality prompts consistent with the semantic and morphological characteristics of microstructures without incurring the costs of large-scale parameter tuning and data annotation, and further effectively improve the segmentation performance under few-shot conditions, has become an urgent problem to be addressed.

To address the aforementioned issues, this paper presents MaterialAlphaSAM, a lightweight domain adaptation framework based on SAM, for the fine segmentation and quantitative characterization of the primary $\alpha$ phase in high-magnification microstructure images. This method incorporates a Self-Prompt module while preserving the strong

representational capacity of general visual foundation models, which adaptively predicts key points from raw microstructure images and furnishes SAM with more accurate and semantically consistent target localization information, thus achieving markedly superior performance to the conventional Grid Prompt strategy under few-shot conditions.

## 2. Materials and Methods

The architecture of the proposed MaterialAlphaSAM is illustrated in Figure 1, which comprises two core functional modules: the Geometry-Constrained Prompt Prior (GCPP) module and the Domain-Adaptation Adapter (DAA) module.

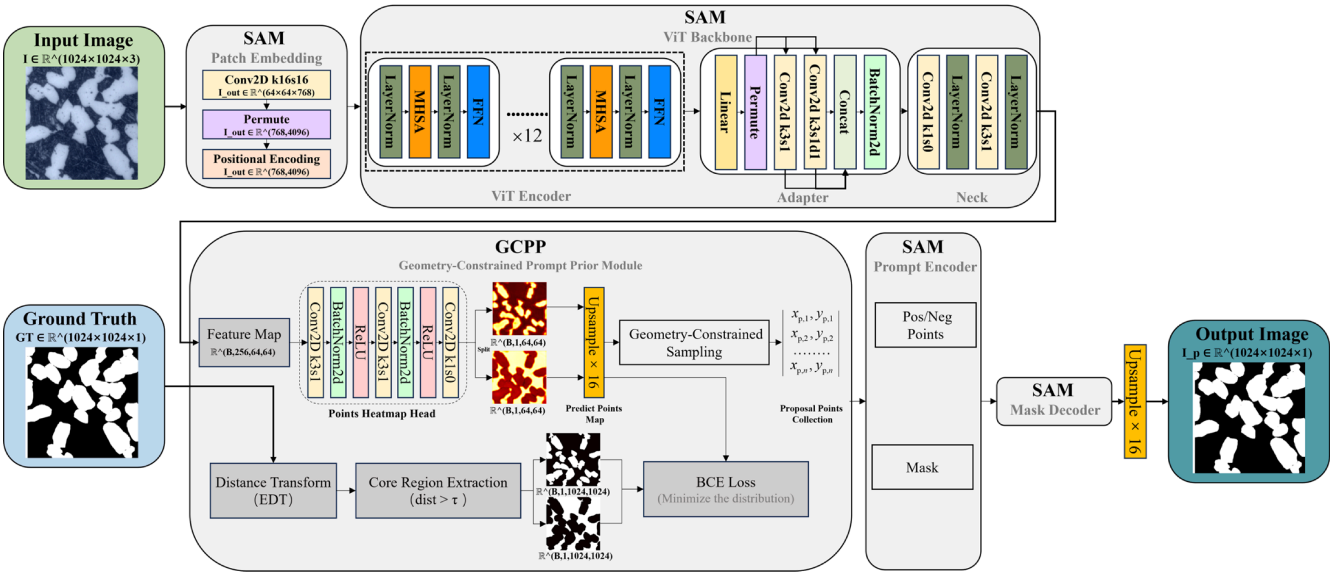



**Figure 1.** Schematic diagram of the MaterialAlphaSAM model.

### *2.1. Geometry-Constrained Prompt Prior Module*

In metallographic microstructure images, the primary α phase typically exhibits a highly discrete size distribution and inhomogeneous spatial characteristics, with the coexistence of fine particles and coarse equiaxed grains and the sparse, random distribution of such targets within the field of view. In few-shot tasks, fixed grid prompt-based strategies struggle to address this characteristic effectively. For one thing, an excessively low grid density tends to result in the omission of numerous fine α phase particles, leading to a notable decline in the recall rate of small-scale targets. For another, blind increases in grid density will introduce a large number of invalid prompt points located in background regions or areas with blurred boundaries. Such semantically ambiguous prompts fail to provide effective supervision; instead, they interfere with the model's discrimination of actual target regions and impair the focusing capability of the Mask Decoder. Especially in lightweight adaptation and few-shot scenarios with extremely limited annotations, the degradation of prompt quality directly amplifies gradient noise, which severely restricts the rapid convergence of the model and its final segmentation performance. To tackle this problem, the GCPP module was designed in this paper, which adaptively predicted the confidence map of foreground and background prompt point pairs from image features and integrated an iterative greedy sampling strategy to generate high-quality point prompts with spatial dispersion and semantic consistency.

Specifically, the GCPP module comprised a lightweight prediction head and an iterative diverse sampler. The prediction head took the feature map $F_m \in \mathrm{R}^{B\times 256\times 64\times 64}$ output by the SAM image encoder as input, and mapped it to a 2-channel feature map $F_{\text{logit}} \in \mathrm{R}^{B\times 2\times 64\times 64}$ through three convolutional layers, which represented the raw con-

fidence maps of foreground and background prompt point pairs, and its mathematical formulation was given as follows.

$$F_{\text{logit}} = \mathrm{H}(F_{\text{sam}}) \tag{1}$$

where $\mathrm{H}(\cdot)$ denotes the prediction head.

The obtained $F_{\text{logit}}$ then underwent channel splitting to yield the confidence maps of the foreground and background prompt point pairs, and was restored to the original image resolution via bilinear upsampling, generating $F_{\text{fg}} \in \mathrm{R}^{B\times1\times1024\times1024}$ and $F_{\text{bg}} \in \mathrm{R}^{B\times1\times1024\times1024}$. The iterative greedy sampling algorithm was adopted for the foreground and background, which selected prompt points at the current highest response positions and implemented circular suppression with a minimum distance threshold to ensure the spatial dispersion of the sampled points and avoid information redundancy and repeated supervision. Assuming that $N^{\text{pos}}$ foreground points were selected, its mathematical formulation could be expressed as follows by taking the sampling process of foreground points as an example. Initializing $R^{(0)} = F_{\text{fg}} \in \mathrm{R}^{B\times1\times1024\times1024}$, with each iteration corresponding to $k = 1, \ldots, N^{\text{pos}}$

$$(x_k, y_k) = \underset{(x.y)}{\operatorname{argmax}} R^{(k-1)}(x, y) \tag{2}$$

$$R^{(k)} = R^{(k-1)} \odot f_{\text{sign}}(d((x, y), (x_k, y_k)) > d_{min} \tag{3}$$

$$f_{\text{sign}}(c) = \begin{cases} 1, & c = \text{true} \\ 0, & c = \text{flase} \end{cases} \tag{4}$$

where $d(\cdot, \cdot)$ denoted the Euclidean distance.

$$d((x, y), (x', y')) = \sqrt{(x - x')^2 + (y - y')^2} \tag{5}$$

If the set was not fully populated after $N_{\text{pos}}$ iterations, the remaining positions with the maximum response were selected in sequence. Finally, the foreground sampling point set $P_{\text{fg}} \in \mathrm{R}^{N_{\text{pos}}\times2}$ was derived.

$$P_{\text{fg}} = \left\{(x_1, y_1), \ldots, (x_{N_{\text{pos}}}, y_{N_{\text{pos}}})\right\} \tag{6}$$

Similarly, the background sampling point set $P_{\text{fg}} \in \mathrm{R}^{N_{\text{pos}}\times2}$ was finally obtained.

$$P_{\text{bg}} = \left\{(x_1, y_1), \ldots, (x_{N_{\text{pos}}}, y_{N_{\text{pos}}})\right\} \tag{7}$$

where $N_{\text{neg}}$ denoted the number of background points selected.

To further enhance the stability of prompt points, the module generated a low-resolution mask input $F_{\text{mask}} \in \mathrm{R}^{B\times1\times256\times256}$. First, threshold segmentation was performed on $F_{\text{fg}}$ to yield a binary foreground mask, which was then downsampled to the size of $256 \times 256$.

$$F_{\text{mask}}(x, y) = f_{\text{interpolate}}(f_{\text{sign}}(F_{\text{fg}}(x, y) > 0)) \tag{8}$$

The mask prompt $F_{\text{mask}}$ and the sampled point prompts $P_{\text{fg}}$ and $P_{\text{bg}}$ were jointly fed into the Prompt Encoder, which generated two types of embeddings including sparse and dense ones, thus furnishing the subsequent Mask Decoder with enhanced semantic priors. Its mathematical formulation was expressed as the following equation.

$$(E_{\text{sparse}}, E_{\text{dense}}) = f_{\text{prompt}}((P_{\text{fg}}, P_{\text{bg}}), F_{\text{mask}}) \tag{9}$$

where $f_{\text{prompt}}(\cdot)$ denoted the Prompt Encoder module of SAM; $E_{\text{sparse}}$ and $E_{\text{dense}}$ represented the encoded sparse and dense embedding vectors, respectively.

Finally, to ensure that the adaptively generated prompts stably targeted the semantic core regions of the primary $\alpha$ phase, this paper further introduced an auxiliary supervision branch based on the distance transform. The branch took the ground truth mask as input, and depicted the spatial distance distribution $d(\cdot)$ from the foreground and background prompt point pairs to the primary $\alpha$ phase boundaries via the Euclidean Distance Transform (EDT). Based on this distribution, the foreground core region $F_{\text{core}_{\text{fg}}}$ and the background core region $F_{\text{core}_{\text{bg}}}$ with high semantic certainty were extracted.

$$\begin{cases} d_{\text{fg}}(p) = \min\limits_{q \in F_{\text{bg}}} \|p - q\|_2 \\ d_{\text{bg}}(p) = \min\limits_{q \in F_{\text{fg}}} \|p - q\|_2 \end{cases} \tag{10}$$

$$\begin{cases} F_{\text{core}_{\text{fg}}} = \left\{ p \middle| d_{\text{fg}}(p) > \tau_{\text{fg}} \right\} \\ F_{\text{core}_{\text{bg}}} = \left\{ p \middle| d_{\text{bg}}(p) > \tau_{\text{bg}} \right\} \end{cases} \tag{11}$$

$$\begin{cases} \tau_{\text{fg}} = \max(\min\limits_{p \in F_{\text{core}_{\text{fg}}}} d_{\text{fg}}(p), 10) \\ \tau_{\text{bg}} = \max(\min\limits_{p \in F_{\text{core}_{\text{bg}}}} d_{\text{bg}}(p), 10) \end{cases} \tag{12}$$

In high-magnification microstructure images, the phase boundaries between the primary $\alpha$ phase and the $\beta$ matrix exhibited characteristics of gradual transition and low contrast, which introduces inherent ambiguity in the physical meaning of the pixels adjacent to the boundaries even though such pixels are rigidly labeled as foreground or background in the ground truth mask. Direct application of pixel-level supervision to these blurred regions forces the model to output extreme positive and negative logits, thereby generating prompt points with semantic ambiguity and spatial instability. In few-shot scenarios with extremely limited annotated samples, this issue would further amplify gradient noise, thus severely restricting the convergence behavior and final segmentation performance of the prompt generation module.

To alleviate the aforementioned issues, the core region is formulated as a morphology-inspired adaptive erosion, in which only stable interior pixels with sufficient thickness and adequate separation from phase boundaries are preserved. Based on this core region, a Binary Cross-Entropy loss with logits (BCEWithLogits) is imposed on the predicted foreground/background prompt confidence maps. This loss explicitly constrains the network to produce high responses within semantically reliable core regions while suppressing ambiguous responses in non-core and transition regions, thereby enhancing the semantic consistency and spatial representativeness of the prompt points.

$$L_{\text{prompt}}^{\text{fg}} = f_{\text{bce}}(F_{\text{fg}}, F_{\text{core}_{\text{fg}}}) \tag{13}$$

$$L_{\text{prompt}}^{\text{bg}} = f_{\text{bce}}(F_{\text{bg}}, F_{\text{core}_{\text{bg}}}) \tag{14}$$

This auxiliary loss acts on the prompt generation module via backpropagation, enabling it to gradually learn prompt patterns that align with the semantics of microstructures and exhibit more stable spatial distribution, thus forming an effective closed-loop supervision mechanism under the condition of limited annotated samples. Benefiting from the geometric priors provided by distance transform, this design drastically reduces the generation of semantically ambiguous or redundant prompt points, enhanced the spatial representativeness and semantic discriminability of prompts, and accelerated the model

convergence process in few-shot scenarios, ultimately improving the stability and quantitative reliability of the segmentation results for the primary α phase.

### *2.2. Domain-Adaptation Adapter Module*

To effectively mitigate the significant domain shift problem of the SAM in the field of material microstructure analysis, this paper introduces a lightweight feature adaptation module after the ViT encoder of the SAM. This module is inserted as a residual connection and performed targeted modulation and compensation on intermediate features with minimal introduction of parameters and computational overhead, thus preserving the pre-trained general visual representation capability of the SAM to the greatest extent while incorporating the prior knowledge of multi-scale textures and low-contrast structures unique to material microstructures.

Specifically, the adaptation module first performed channel compression on the output features of the ViT via linear projection, mapping the original 768-dimensional features to a 128-dimensional bottleneck space. This approach drastically reduced the computational complexity of subsequent operations while ensuring the sufficient preservation of discriminative information. A dual-scale depthwise separable convolution branch was subsequently incorporated into this low-dimensional feature space to enable multi-scale feature modeling. Herein, the small-scale branch focused on extracting high-frequency detailed information including fine equiaxed grains, local rough boundaries and micro-texture intersections. The large-scale branch adopts a 3 × 3 depthwise separable convolution with a dilation rate of 2 (dilation = 2, padding = 2) to expand the receptive field, thereby capturing the overall morphological structures and distribution characteristics of low-contrast regions of the primary α phase more effectively. Its mathematical formulation is given as follows.

$$F_{\text{down}} = f_{\text{linear}}(F_{\text{vit}}) \tag{15}$$

$$F_{\text{small}} = f_{\text{DWConv}_{3\times3}}(F_{\text{down}}) \tag{16}$$

$$F_{\text{large}} = f_{\text{DWConv}_{3\times3}^{\text{dil}=2}}(F_{\text{down}}) \tag{17}$$

where $F_{\text{vit}} \in \mathrm{R}^{B\times768\times64\times64}$ denoted the output of the ViT Backbone of SAM; $F_{\text{down}} \in \mathrm{R}^{B\times128\times64\times64}$ represented the compressed feature; $F_{\text{small}} \in \mathrm{R}^{B\times128\times64\times64}$ and $F_{\text{large}} \in \mathrm{R}^{B\times128\times64\times64}$ denoted the multi-scale features respectively.

The features from the two scale branches were subsequently concatenated along the channel dimension, and then processed sequentially through Batch Normalization and the GELU activation function to achieve normalization and non-linear mapping. Finally, a 1 × 1 convolution was employed to restore the feature dimension to the original channel number, thus enabling residual fusion with the backbone features.

$$F_{\text{fuse}} = \text{Concat}(F_{\text{small}}, \mathrm{F}_{\text{large}}) \tag{18}$$

$$F_{\text{act}} = f_{\text{act}}(f_{\text{BN}}(F_{\text{fuse}})) \tag{19}$$

$$F_{\text{adapter}} = f_{\text{up}}(F_{\text{act}}) \tag{20}$$

Through the aforementioned dual-scale feature modeling and channel-adaptive remapping mechanism, the adaptation module enabled the model to effectively enhance its representational capability for the multi-scale distribution characteristics of the primary α phase in material microstructure images, and perform targeted feature compensation for low-contrast boundary regions, thus drastically narrowing the representational gap between general visual priors and domain-specific structural features of materials.

### *2.3. Loss Function*

To achieve high accuracy and boundary stability for the fine segmentation of the primary α phase, this paper constructed a composite loss function. Via the organic integration of intersection-over-union loss, pixel-level confidence loss and boundary-aware loss, the function realized comprehensive supervision of the model under low-contrast and multi-scale conditions.

This loss function consisted primarily of four components, including the Dice loss, the Focal loss, the Boundary loss and the prompt supervision loss mentioned in Section 2.1.

$$L_{\text{total}} = L_{\text{focal}} + L_{\text{dice}} + 0.4 \cdot L_{\text{boundary}} + 0.6 \cdot (L_{\text{prompt}}^{\text{fg}} + L_{\text{prompt}}^{\text{bg}}) \tag{21}$$

## 3. Results

### *3.1. Experimental Setup*

The experimental workstation was equipped with an Intel Xeon E5-2650 processor, 376 GB of RAM and three NVIDIA L4 24 G graphics cards. For the deep learning framework, PyTorch 1.8 was adopted, along with the NVIDIA CUDA 11.2 GPU runtime platform and the cuDNN 8.0 deep learning GPU acceleration library.

The experimental dataset was a self-constructed dataset of titanium alloy microstructure images with a resolution of 2048 × 2048. The dataset was cropped to a resolution of 1024 × 1024, generating a total of 50 local microstructure images, and then subjected to dataset division at the ratio of 6:3:1. The resulting training set contained 30 images, the validation set 15 images and the test set five images. In the training process of the Material Alpha SAM, titanium alloy microstructure images cropped to 1024 × 1024 were used as input, and data augmentation for metallographic microstructure images was realized by adopting the strategies of random horizontal flipping and normalization. In the experiments, the network was set with the training parameters of a batch size of two, 10 training epochs and an initial learning rate of $1 \times 10^{-3}$, and the AdamW optimizer was employed for network optimization. During the network training, a cosine learning rate adjustment strategy was adopted with an adjustment cycle of 10 epochs, and the minimum learning rate was set to $1 \times 10^{-5}$. Meanwhile, to alleviate the overfitting problem, the backbone network of the MaterialAlphaSAM was loaded with the smaller pre-trained weights of ViT-B in the experiments.

### *3.2. Experimental Evaluation Metrics*

To verify the effectiveness of the algorithm proposed in this paper, the intersection over union (IoU) and F1-score were adopted as evaluation metrics to conduct a comparative assessment of the algorithm's effectiveness. The definition formulas for each metric were given as follows.

$$f_{\text{IoU}} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{22}$$

$$f_{F1-\text{Score}} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{23}$$

where TP represented the total number of correctly segmented pixels for the primary α phase in the segmentation results of the microstructure images; TN represented the total number of correctly segmented pixels for the background in the segmentation results of the microstructure images; FP represented the total number of incorrectly segmented pixels for the primary α phase in the segmentation results of the microstructure images; and FN represented the total number of incorrectly segmented pixels for the background in the segmentation results of the microstructure images.

### 3.3. Ablation Experiments

To verify the effectiveness of the proposed geometric constraint prompt prior module and domain adaptation module, ablation experiments were conducted on the titanium alloy microstructure dataset. The ablation experiments performed comparative analysis based on metrics including the IoU and F1-score of the segmentation results. In all the experiments, consistent data division, parameter scale and training strategies were adopted for all the model variants, and differentiated settings were only implemented in terms of the network structure or encoder freezing strategy. The semantic segmentation model proposed in Reference [11] was employed as the baseline model, which adopted a structure composed of the original SAM, a fully frozen encoder and fine-tuning of the MLP.

#### 3.3.1. Module-Level Ablation Experiment Results and Analysis

Table 1 presents the descriptions of the structural configurations for all the network variants of the Material Alpha SAM, and Table 2 summarizes the corresponding ablation experiment results. Table 2 shows that Baseline + GCPP outperforms the Baseline by 3.87% in the IoU and 2.55% in the F1-Score, Baseline + DAA by 3.73% in the IoU and 2.44% in the F1-Score, and Baseline + GCPP + DAA by 5.3% in the IoU and 3.77% in the F1-Score. The results demonstrate that both the GCPP module and the DAA module can independently enhance the model performance with a fully frozen encoder, and the two modules exhibit a distinct complementary gain effect. GCPP effectively boosts the spatial attention of the model to the core regions and boundary contours of the primary $\alpha$ phase by introducing structural geometric priors. The DAA mitigates the domain shift between pre-trained representations of natural images and the texture distribution of microstructures through domain alignment of intermediate features.

**Table 1.** Variant structure of the MaterialAlphaSAM.

| Network Variant Module | Description |
|---|---|
| Baseline | Original SAM + backbone freezing + MLP fine-tuning |
| Baseline + GCPP | Original SAM + backbone freezing + MLP fine-tuning + GCPP module |
| Baseline + DAA | Original SAM + backbone freezing + MLP fine-tuning + DAA module |
| Baseline + GCPP + DAA | Original SAM + backbone freezing + MLP fine-tuning + GCPP module + DAA module |

**Table 2.** Ablation experimental results.

| Network Variant Module | IoU (%) | F1-Score (%) |
|---|---|---|
| Baseline | 82.9 | 90.1 |
| Baseline + GCPP | 86.1 | 92.4 |
| Baseline + DAA | 86.0 | 92.3 |
| Baseline + GCPP + DAA | 87.3 | 93.5 |

#### 3.3.2. Analysis of the Influence of Encoder Unfreezing Strategies on Model Performance

From the aforementioned ablation experiments, it can be concluded that introducing the GCPP module and the DAA module can significantly enhance the ability of the model to focus on the regions and boundaries of primary $\alpha$ phases when the encoder is fully frozen. However, the pre-training of the SAM is mainly based on natural images, and the representations learned by its encoder show significant differences from the texture, scale and structural features of metallic microstructures. Under few-shot conditions, such domain shift can further limit the model performance, making the strategy that only relies

on backbone freezing and lightweight fine-tuning unable to fully adapt to downstream tasks. Therefore, this paper further designs multiple hierarchical unfreezing strategies for the encoder and analyzes the complementary effects and gain performances of the GCPP and DAA modules under different unfreezing settings. The specific unfreezing configurations are presented in Table 3, and the comparison results are shown in Tables 4–7.

**Table 3.** Configuration specification tables for different frozen strategies.

| No. | Frozen Strategy | Description |
|---|---|---|
| E0 | Fully Frozen | The ViT encoder is fully frozen |
| E1 | Last Block | Only the last Transformer Block is unfrozen |
| E2 | Last 2 Blocks | The last two Transformer Blocks are unfrozen |
| E3 | Last 3 Blocks | The last three Transformer Blocks are unfrozen |
| E4 | Last 4 Blocks | The last four Transformer Blocks are unfrozen |
| E5 | Last 5 Blocks | The last five Transformer Blocks are unfrozen |

**Table 4.** Comparison of different frozen strategies for Baseline.

| Frozen Strategy | IoU (%) | F1-Score (%) | Trainable Param (%) |
|---|---|---|---|
| E0 | 82.90 | 90.10 | 1.20 |
| E1 | 86.57 | 92.59 | 8.54 |
| E2 | 86.75 | 92.64 | 16.06 |
| E3 | 85.78 | 91.79 | 23.58 |
| E4 | 85.12 | 91.26 | 30.93 |
| E5 | 83.47 | 90.88 | 38.63 |

**Table 5.** Comparison of different frozen strategies for Baseline + DAA.

| Frozen Strategy | IoU (%) | F1-Score (%) | Trainable Param (%) |
|---|---|---|---|
| E0 | 86.00 | 92.30 | 1.30 |
| E1 | 87.20 | 93.00 | 8.64 |
| E2 | 86.69 | 92.60 | 16.16 |
| E3 | 86.68 | 92.64 | 23.68 |
| E4 | 85.74 | 91.99 | 31.03 |
| E5 | 85.02 | 91.60 | 38.73 |

**Table 6.** Comparison of different frozen strategies for Baseline + GCPP.

| Frozen Strategy | IoU (%) | F1-Score (%) | Trainable Param (%) |
|---|---|---|---|
| E0 | 86.10 | 92.40 | 1.20 |
| E1 | 87.81 | 93.40 | 8.54 |
| E2 | 88.80 | 94.00 | 16.06 |
| E3 | 89.10 | 94.10 | 23.68 |
| E4 | 87.95 | 93.58 | 31.03 |
| E5 | 87.12 | 92.89 | 38.73 |

The ablation experiment results demonstrate that for the Baseline model, the model performance improves continuously as the number of unfrozen encoder layers increases gradually from E0 to E2, which indicates that moderate unfreezing of upper Transformer Blocks helps alleviate the domain shift between pre-trained features and downstream tasks. However, the model performance declines significantly when the unfreezing depth

is further extended to E3 and beyond, which reveals that an excessive number of trainable parameters is prone to induce overfitting under the condition of limited training samples, thereby weakening the stability of pre-trained representations.

**Table 7.** Comparison of different frozen strategies for Baseline + GCPP + DAA.

| Frozen Strategy | IoU (%) | F1-Score (%) | Trainable Param (%) |
|---|---|---|---|
| E0 | 87.30 | 93.50 | 1.30 |
| E1 | 88.89 | 94.00 | 8.74 |
| E2 | 88.79 | 94.23 | 16.26 |
| E3 | 89.53 | 94.40 | 23.78 |
| E4 | 87.40 | 93.17 | 31.13 |
| E5 | 86.00 | 92.16 | 38.83 |

With the introduction of the DAA module, the model achieves optimal performance under shallow unfreezing strategies such as E1, and the fluctuation of its overall performance with the variation in unfreezing depth is significantly smaller than that of the Baseline, demonstrating stronger robustness. Nevertheless, the performance still degrades to a certain extent under the moderate unfreezing depth of E3, which indicates that although the DAA module can effectively alleviate the inter-domain shift in the feature distribution, its ability to suppress overfitting remains insufficient when the encoder is excessively unfrozen.

In contrast, Baseline + GCPP not only maintains high segmentation performance at the moderate unfreezing depth of E3 but also achieves a further performance improvement, and is significantly superior to both Baseline and Baseline + DAA. This phenomenon indicates that the geometric constraint prompt prior exhibits a regularization effect in guiding the feature learning process, effectively enhancing the ability of the model to focus on the key information of the primary α phase microstructure. Thus, it suppresses overfitting while improving the plasticity of model.

When the GCPP and DAA modules are jointly employed, the model achieves the overall optimal segmentation performance under the E3 unfreezing strategy and attains a favorable balance between accuracy and the scale of trainable parameters. The geometric prior guidance and domain adaptation mechanism yield a significant synergistic gain effect.

#### 3.3.3. Analysis of the Influence of Grid Prompts with Different Sizes on Model Performance

To further analyze the influence of the prompt design strategy on SAM-based microstructure segmentation performance and verify the applicability and limitations of the conventional Grid Prompt in high-magnification metallographic image scenarios, this paper designs ablation experiments under different Grid Prompt configurations. Grid Prompt serves as the most commonly used unsupervised prompting method in SAM, and its core hypothesis is that potential target regions are covered through regularized spatial sampling, thereby guiding the model to accomplish segmentation. However, in the microstructure images of primary α phases, target regions typically exhibit the characteristics of irregular morphologies, significant scale differences and highly inhomogeneous spatial distributions. Fixed grid sampling fails to effectively balance target integrity and background interference suppression, which may cause prompt points to fall on phase boundaries, background or noise regions, thus weakening the effective response of the model.

Based on this, the influence of different grid sampling densities on segmentation performance is mainly investigated, and a comparative analysis is conducted between

them and the GCPP strategy proposed in this paper. The SAM unfreezing strategy is fixed as E2, the setting for Grid Prompt to achieve optimal performance.

On the basis of the experimental results in Table 8, Grid Prompt, the most widely adopted unsupervised prompting strategy in the SAM, has particularly prominent inherent limitations in the microstructure segmentation task of high-magnification metallographic images. The target regions in the microstructure images of the primary α phase usually exhibit irregular morphologies, significant scale variations and low contrast. These characteristics make fixed grid sampling unable to cover target regions effectively and easily locate prompt points in irrelevant regions such as the background and phase boundaries, thus impairing the segmentation accuracy of the model. At low sampling densities, such as those of Grid 4 and Grid 8, Grid Prompt still maintains relatively stable performance; however, with increasing grid density, such as that of Grid 128, the performance improvement gradually plateaus. This phenomenon indicates insufficient adaptability of the fixed grid sampling method in the face of complex target scenarios. In contrast, the GCPP strategy proposed in this paper effectively addresses the local noise problems that may arise in complex images with Grid Prompt by adaptively generating prompt points, and it can capture the details of target regions more accurately, especially in the fine particles and blurred phase boundary regions of the primary α phase. According to the experimental results in Table 7, the GCPP strategy achieves remarkable improvements in key metrics, including the IoU and F1-score, and performs exceptionally well in fine-grained segmentation tasks in particular. In summary, the GCPP strategy not only overcomes the inherent limitations of Grid Prompt but also significantly enhances the accuracy and robustness of segmentation through the adaptive generation of prompt points, which furnishes a more flexible and efficient solution for metallographic image analysis tasks.

**Table 8.** Comparison results of different grid sampling densities.

| Grid Sampling Density | IoU (%) | F1-Score (%) |
|---|---|---|
| 4 | 86.74 | 92.62 |
| 8 | 86.68 | 92.59 |
| 16 | 86.73 | 92.63 |
| 32 | 86.75 | 92.64 |
| 64 | 86.73 | 92.62 |
| 128 | 86.81 | 92.67 |

### *3.4. Comparative Experiments*

To verify the effectiveness of the proposed MaterialAlphaSAM, segmentation comparison experiments are conducted with several mainstream advanced semantic segmentation networks on the metal microstructure dataset. The comparative methods include FCN [14], Unet [15], SegNet [16], DeeplabV3 [17], PSPNet [18] and SAM Baseline. Table 9 presents the quantitative results of comparative experiments on the titanium alloy microstructure dataset, and Figure 2 shows partial segmentation results of various comparative methods on the metallographic dataset. In the figures, the first column, Raw Image, from left to right represents the input titanium alloy microstructure images for testing and the second column, Ground Truth, denotes the real label images. The third to eighth columns display the semantic segmentation results of the FCN, UNet, SegNet, DeepLabV3, PSPNet and SAM Baseline algorithms, respectively. The ninth column OURS illustrates the semantic segmentation results of the MaterialAlphaSAM proposed in this paper and the tenth column Generated Prompt Points indicates the prompt points automatically generated during the training process of the GCPP module proposed in this paper.

**Table 9.** Quantitative comparison results of the titanium alloy microstructure datasets.

| Method | IoU (%) | F1-Score (%) |
|---|---|---|
| FCN | 73.56 | 82.65 |
| UNet | 80.10 | 87.74 |
| SegNet | 79.46 | 87.77 |
| DeepLabV3 | 81.12 | 88.79 |
| PSPNet | 81.24 | 88.97 |
| SAM Baseline | 86.75 | 92.64 |
| MaterialAlphaSAM | 89.53 | 94.40 |

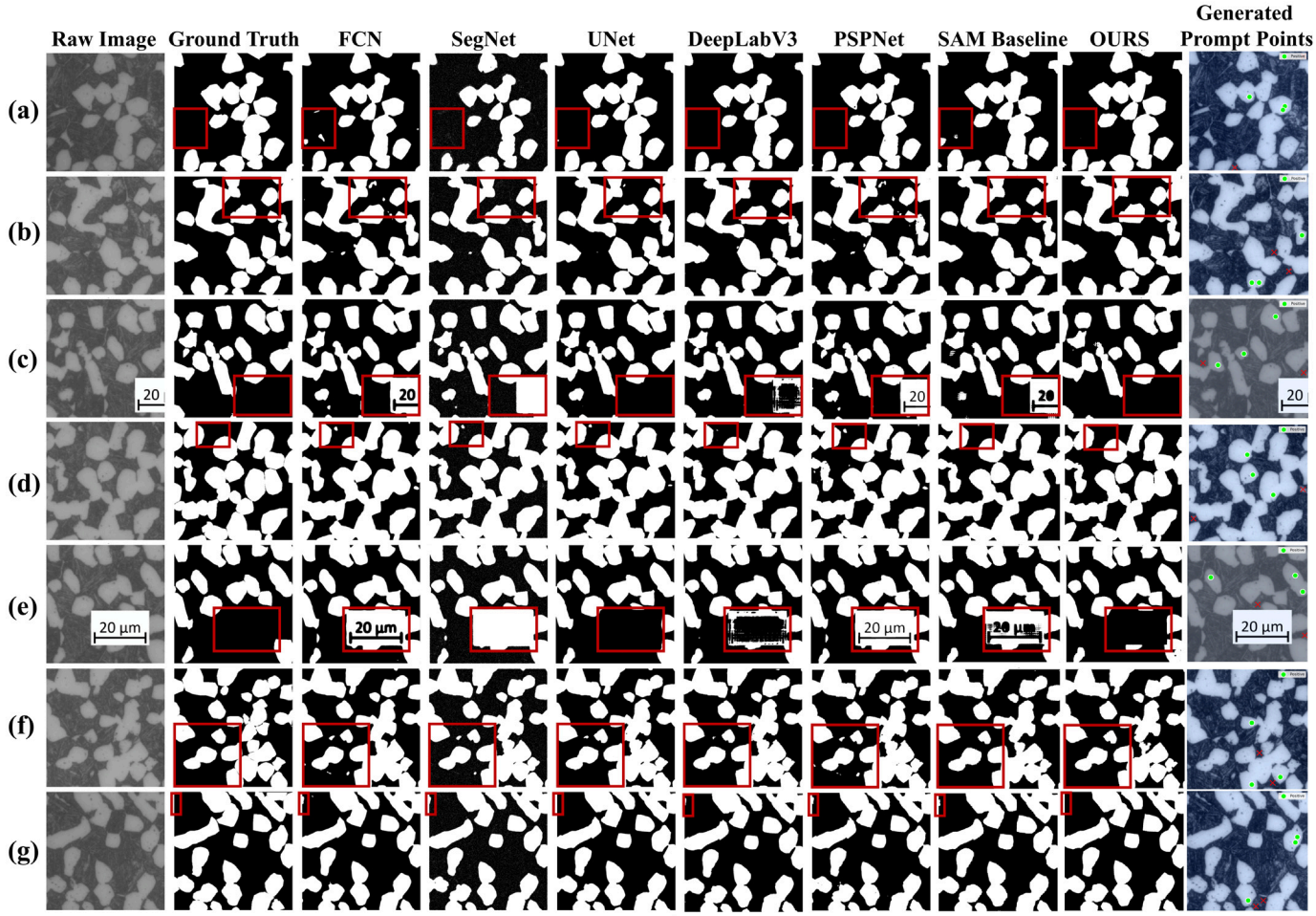



**Figure 2.** Qualitative comparison results of titanium alloy microstructure dataset.

Quantitative and qualitative analyses are performed separately. The quantitative evaluation metrics in Table 9 demonstrate that the MaterialAlphaSAM achieves the optimal performance in the segmentation task of titanium alloy microstructures. Specifically, traditional convolutional networks such as the FCN yield the weakest performance with an IoU of 73.56% and an F1-score of 82.65%, which is mainly restricted by insufficient local receptive fields and inadequate capability of global context modeling. UNet and SegNet show improved performance in local structure capture with IoUs of 80.10% and 79.46%, respectively. However, their performance is still lower than that of the MaterialAlphaSAM. DeepLabV3 and PSPNet which adopt atrous convolution or pyramid pooling, can enhance global information perception, achieving IoUs of 81.12% and 81.24% as well as F1-scores of 88.79% and 88.97%, respectively. Compared with that of the FCN, their performance is significantly improved but still notably lower than that of the MaterialAlphaSAM. On the basis of the configuration of the ablation experiments in Section 3.3, the SAM Base-

line achieves an IoU of 86.75% and an F1-score of 92.64%, which reflects its advantages in global structure modeling. However, the MaterialAlphaSAM proposed in this paper further integrates domain adaptation and an adaptive prompting mechanism, leading to remarkable improvements in the segmentation of local details and weak boundaries with an IoU of 89.53% and an F1-score of 94.40%. This represents an increase of 3.2% and 1.90% compared with SAM Baseline in terms of the IoU and F1-score, respectively. Overall, the MaterialAlphaSAM achieves the highest IoU and F1-score among all comparative methods, which verifies that the proposed modules can significantly improve the performance of titanium alloy microstructure segmentation under low-sample scenarios.

Figure 2a–g present row-wise the comparative segmentation results of the proposed MaterialAlphaSAM and various mainstream semantic segmentation methods under complex scenarios of the primary α phase microstructure, including highly uneven scale distribution, occlusion, complex background textures and low-contrast boundaries. Overall, in comparison with the FCN, UNet, SegNet, DeepLabV3, PSPNet and SAM Baseline, the Material Alpha SAM can more effectively fuse local geometric detail information and global semantic prior to the microstructure, enabling the model to maintain high segmentation accuracy and stability when confronted with blurred α phase boundaries, drastic scale variations and background noise interference.

In cases with complex background textures and low local contrast, as illustrated in Figure 2b, traditional CNN method tends to misclassify partial β matrix textures as the α phase within the red annotated regions, whereas the segmentation results of the SAM Baseline are highly sensitive to the positions of prompt points. In contrast, the method proposed in this paper can more accurately focus on the core regions and real boundary morphologies of the α phase, thus achieving fine segmentation of α phase boundaries in low-contrast regions. In complex scenarios including local adhesion of α phases, dense distributions of small-scale α phases and irregular morphologies, as shown in Figure 2a,d–g, FCN, UNet and other comparable methods generally suffer from over-segmentation or missing detection, and DeepLabV3 as well as PSPNet show insufficient responsiveness to small-scale targets. The MaterialAlphaSAM can effectively separate adjacent α phases and preserve fine microstructures while maintaining the integrity of large-scale α phases, which significantly enhances the geometric consistency of the segmentation results.

In addition, as presented in Figure 2c,e, some samples exhibit occlusion and interference caused by magnification markers. Compared with the FCN, SegNet, DeepLabV3, PSPNet and SAM Baseline, these models generally misclassify magnification markers as the primary α phase within such regions, leading to obvious false detection. In contrast, the MaterialAlphaSAM proposed in this paper can effectively distinguish non-microstructure semantic information from actual microstructural features and exhibits a stronger suppression capability for regions with magnification markers, with its segmentation results reflecting the morphological characteristics of real microstructures more accurately. In summary, the qualitative comparison results on the metallographic microstructure dataset demonstrate that the proposed MaterialAlphaSAM based on adaptive prompt generation and domain adaptation can effectively improve the segmentation stability and refinement of the SAM in complex microstructure scenarios, and it has remarkable advantages in the accurate segmentation of the primary α phase.

## 4. Discussion

### *4.1. Comparison with State-of-the-Art Methods*

This paper proposes a MaterialAlphaSAM model based on the GCPP module and the DAA module, which is designed to improve the segmentation accuracy of microstructure images especially those of the primary α phase. Ablation experiments conducted on the

titanium alloy microstructure dataset further verify the independence and complementary gain effect of the GCPP and DAA modules, demonstrating significant performance improvements. Specifically, the GCPP enhances the spatial attention capability of the model to target regions and boundary contours, while the DAA effectively improves the segmentation accuracy and stability of microstructure images by alleviating the domain shift problem. By integrating the effects of these two modules, the model achieves superior performance over baseline methods on multiple performance metrics including the IoU and F1-score.

Compared with existing microstructure segmentation methods, the MaterialAlphaSAM is more adaptable for processing microstructure images with significant scale variations, highly irregular morphologies and low boundary contrast. Traditional convolutional neural network-based methods rely mainly on local texture features and tend to suffer from over-segmentation or missing detection when confronted with complex backgrounds and low-boundary scenarios. While the SAM-based Grid Prompt strategy shows a certain degree of robustness in general scenarios, its underlying assumption that relies on regularized spatial sampling is often invalid in microstructure images. The GCPP strategy proposed in this paper adaptively generates prompt points with geometric semantic directivity, which effectively avoids the introduction of numerous invalid or noisy prompts by fixed grid sampling in complex structural scenarios and significantly enhances the perceptual capability of the model for fine α phase particles and blurred phase boundaries. On this basis, the integration of the DAA module further boosts the domain adaptability of the model, enabling the SAM encoder pre-trained on natural images to better adapt to the texture statistical characteristics and scale distribution laws of microstructure images, thus improving the overall segmentation performance and generalization capability.

### *4.2. Advantages and Contributions of the Proposed Models*

#### 4.2.1. Geometry-Constrained Prompt Prior Module

The GCPP module proposed in this paper essentially incorporates the geometric and morphological priors of microstructures explicitly into the prompt generation process, transforming the distribution of prompt points from uniform coverage to structure-aware driving. The key advantages of this strategy manifest in three core aspects. First, as illustrated by the mask of the predicted foreground point set during the training process in Figure 3, at the spatial localization level, the GCPP guides the preferential distribution of prompt points in the stable core regions and representative structural positions of the α phase through geometric constraints on the foreground regions, markedly reducing the probability of prompt points falling into the background or noisy phase boundary regions. Such structural constraints endow the prompts themselves with a certain discriminative property, thus improving the response quality of the SAM at the source.

Furthermore, from the perspective of feature learning, the GCPP exerts an implicit regularization effect on the model during the training process. In experiments, it can be observed that the Baseline and Baseline + DAA models exhibit performance degradation when the encoder of moderate depth is unfrozen (e.g., E3), whereas the model retains and even further improves its performance after the integration of the GCPP. This finding indicates that the geometrically constrained prompt not only provides a more stable supervision signal but also mitigates the overfitting risk caused by the expansion of the parameter scale to a certain extent.

Finally, from the perspective of task adaptation, the GCPP essentially narrows the semantic gap between the original general segmentation capability of the SAM and the demand for domain-specific fine-grained segmentation, enabling the model to focus more

on the structural regions with genuine materials science significance in microstructures, instead of merely relying on texture or local response intensity.

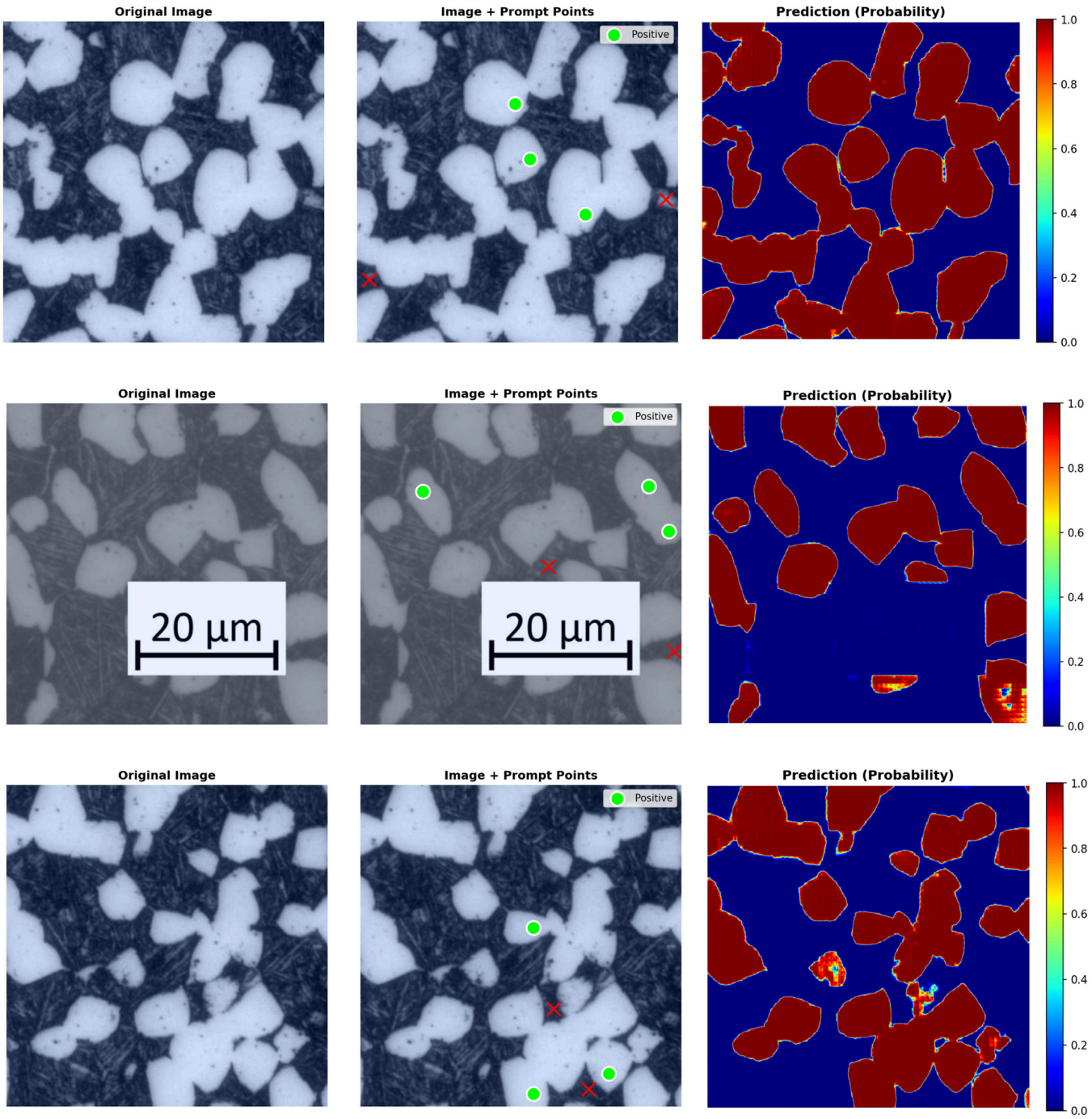



**Figure 3.** Visualization of predicted foreground point set mask and point prompt during training. Green dots represent positive prompts, red crosses represent negative prompts.

#### 4.2.2. Synergistic Relationship Between the Domain-Adaptation Adapter Module and Prompt Guidance

Unlike the constrained prompting of the GCPP, the DAA acts primarily on distribution alignment in the feature representation space. Since the SAM encoder is pre-trained on large-scale natural images, its high-level features exhibit significant discrepancies from those of microstructure images in terms of texture statistics, scale distribution and structural representation, and such domain shift is particularly pronounced under few-shot conditions. Ablation experiments demonstrate that the DAA can significantly improve

model performance under shallow encoder unfreezing strategies and effectively reduce the amplitude of performance fluctuations across different unfreezing depths, which indicates its stable effect in alleviating domain distribution mismatch. However, as the number of unfrozen encoder layers increases further, relying solely on feature alignment still fails to fully suppress the risk of overfitting. When the GCPP is combined with the DAA, the model achieves the overall optimal performance at moderate encoder unfreezing depths.

Overall, by integrating adaptive prompt point generation and the multi-scale feature modeling mechanism, the MaterialAlphaSAM can effectively address the commonly encountered issues of scale inhomogeneity, morphological complexity and background interference in primary $\alpha$ phase microstructures, and yield stable and high-precision segmentation results in complex scenarios.

Despite the excellent performance the MaterialAlphaSAM demonstrates in multiple experimental sets, certain limitations and challenges remain that warrant further investigation. First, at the data level, although the model has achieved promising performance under few-shot conditions on the self-constructed microstructure dataset, microstructure images themselves exhibit strong diversity and complexity. For rare microstructural morphologies or extreme imaging conditions, the full potential of the model can only be unleashed with a more abundant data distribution. Second, at the model training level, as the number of unfrozen encoder layers increases further, although the GCPP and DAA alleviate the overfitting issue to a certain extent, performance degradation may still occur under the constraint of limited sample size. How to further enhance the regularization capability of the model remains a topic worthy of in-depth exploration. Furthermore, in terms of computational efficiency, the introduction of the GCPP increases the computational overhead to a certain extent, particularly in the analysis of high-resolution microstructure images, which imposes higher requirements on the efficiency of training and inference. How to further optimize the computational cost while ensuring segmentation accuracy constitutes a key issue that needs to be focused on in the engineering application of the model. Finally, although the method proposed in this paper has achieved favorable performance in the task of metallographic microstructure segmentation, its generalization capability in more complex cross-domain scenarios such as medical imaging and remote sensing images still needs to be verified through further experiments.

### *4.3. Statistical Validation and Comparative Analysis*

To address the statistical reliability concerns inherent in small-sample evaluation, we adopt a rigorous multi-faceted validation framework comprising 5-fold cross-validation, Bootstrap confidence interval estimation, and Wilcoxon signed-rank significance testing. All competing methods are evaluated under a strictly unified experimental protocol: a fixed set of 50 samples drawn with a deterministic random seed is partitioned into five mutually exclusive folds of equal size, with all methods sharing identical fold assignments and sample indices throughout. Experiments are conducted under two training sample scales ($n$ = 10 and $n$ = 50) to systematically assess model behavior across varying degrees of data scarcity.

#### 4.3.1. Per-Fold Performance Stability Analysis

Figures 4 and 5 illustrate the per-fold performance trends and corresponding 95% Bootstrap confidence intervals of all competing methods across the five evaluation folds under training sample sizes of 10 and 50, respectively. Across both sample scales, MaterialAlphaSAM consistently demonstrates pronounced performance superiority and cross-fold stability. The core segmentation metrics—IoU, F1-Score, and Recall—remain within a persistently high numerical range across all five folds, with inter-fold variation substantially

smaller than that observed in any of the comparative methods. This behavior reflects strong robustness to the stochastic nature of data partitioning, indicating that the proposed method maintains highly consistent performance regardless of the specific data composition encountered in each fold. Moreover, the 95% Bootstrap confidence intervals associated with MaterialAlphaSAM on the core metrics are characteristically narrow, providing further evidence of low estimation uncertainty—a property of particular importance in small-sample evaluation contexts where performance estimates are susceptible to high variance.

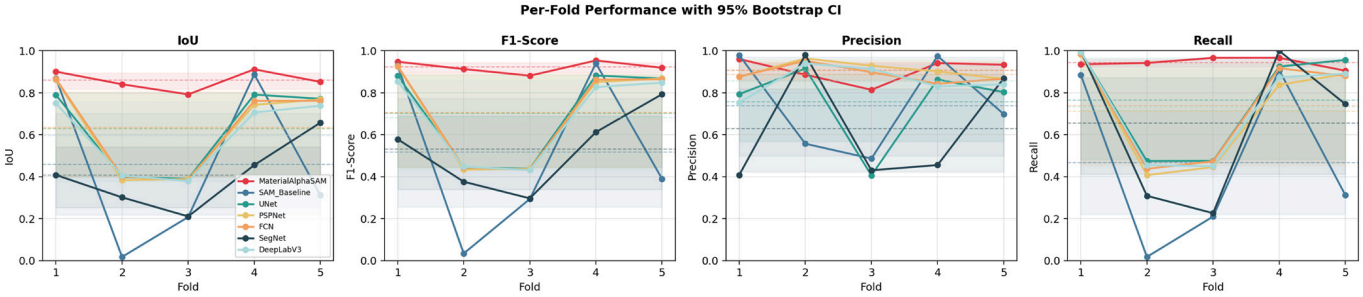



**Figure 4.** Bootstrap confidence intervals (95%) of evaluation metrics under a sample size of 10.

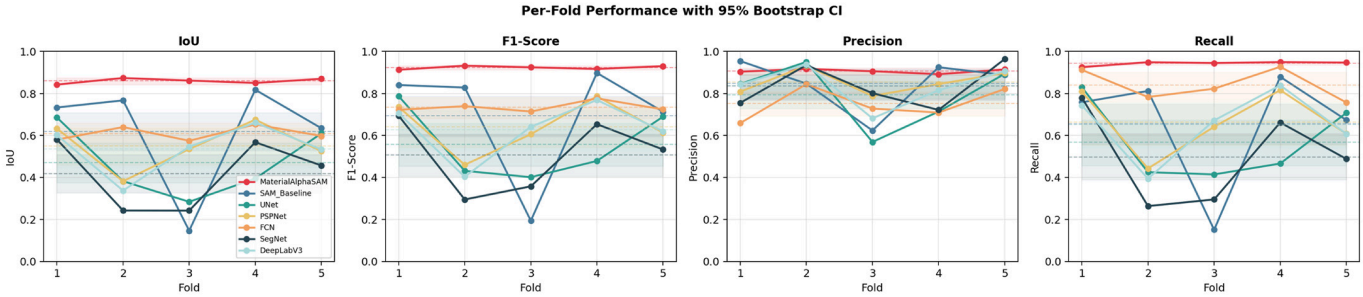



**Figure 5.** Bootstrap confidence intervals (95%) of evaluation metrics under a sample size of 50.

In stark contrast, the vanilla SAM Baseline exhibits severe cross-fold instability. In several evaluation folds, both IoU and F1-Score degrade catastrophically to near-zero values in the presence of challenging samples, a phenomenon that directly corroborates the central problem motivation of this work: the original SAM is inherently vulnerable to distributional shift in domain-specific industrial scenarios. When the feature distribution of the training data diverges substantially from the natural image pre-training domain, the model's segmentation capability undergoes catastrophic collapse. Correspondingly, the Bootstrap confidence intervals of the SAM Baseline span nearly the entire performance range in multiple folds, signaling unacceptably high estimation uncertainty and an inability to provide reliable predictions in deployment-critical applications.

Conventional segmentation architectures including UNet, PSPNet, and FCN exhibit moderate cross-fold fluctuations. While these methods achieve performance approaching that of MaterialAlphaSAM in select folds, they consistently experience varying degrees of degradation across the remaining folds, and their overall cross-fold consistency remains significantly inferior to the proposed method. SegNet additionally exhibits pronounced performance deterioration in individual folds, further underscoring the limited adaptability of classical segmentation architectures to the present industrial few-shot scenario.

#### 4.3.2. Statistical Significance Analysis

Figures 6 and 7 present, in heatmap form, the Wilcoxon signed-rank test $p$-values of each competing method relative to SAM Baseline, providing a quantitative assessment of the statistical significance of the observed performance differences. A threshold of $p < 0.05$

is adopted as the criterion for statistical significance, indicating that the performance gap between the target method and the baseline is statistically meaningful.

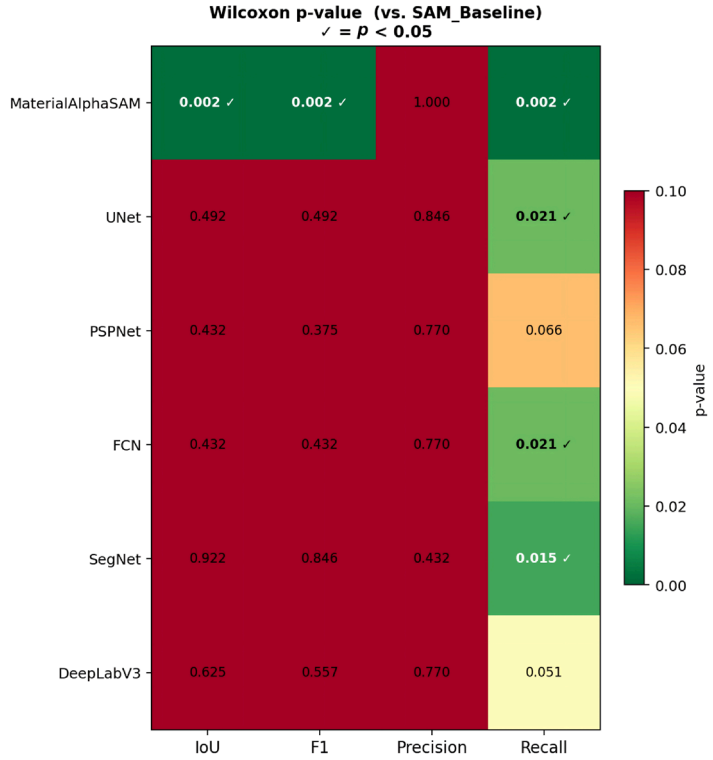



**Figure 6.** *p*-values of the Wilcoxon signed-rank test relative to SAM Baseline under a sample size of 10.

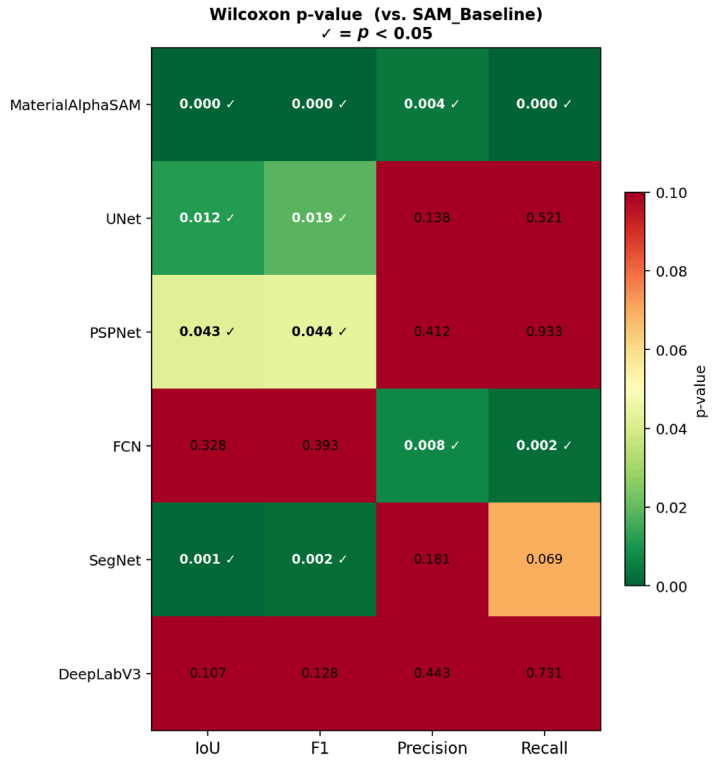



**Figure 7.** *p*-values of the Wilcoxon signed-rank test relative to SAM Baseline under a sample size of 50.

Among the four evaluation metrics, MaterialAlphaSAM achieves statistically significant improvements over the SAM Baseline on the core metrics of IoU and F1-Score

($p < 0.05$), with $p$-values falling well below the conventional significance threshold, thereby demonstrating a high degree of statistical confidence. These results provide compelling evidence that the performance gains of the proposed method over the baseline SAM in terms of overall segmentation quality are not attributable to random variation but are undergirded by a robust statistical foundation. With regard to the Precision metric, the $p$-value for MaterialAlphaSAM relative to SAM Baseline is comparatively large, indicating no statistically significant difference between the two methods in this respect. This finding is nonetheless interpretable: the SAM Baseline already achieves reasonable Precision, and its principal performance bottleneck lies in substantially deficient Recall—i.e., a large proportion of target regions remain undetected—rather than an elevated false-positive rate. The proposed method effectively addresses this limitation by substantially improving Recall, thereby achieving significant gains in IoU and F1-Score without compromising Precision.

Taken holistically, among all comparative methods, MaterialAlphaSAM is the only approach that attains statistically significant improvements on both IoU and F1-Score—the two comprehensive performance metrics—thereby providing robust statistical validation of the proposed framework for the task of industrial microscopic image segmentation.

#### 4.3.3. Overall Performance Analysis Under 5-Fold Cross-Validation

MaterialAlphaSAM surpasses all comparative methods by a substantial margin across all four evaluation metrics. On the two core comprehensive metrics (IoU and F1-Score), the proposed method achieves considerable improvements over all conventional segmentation architectures, with a clearly dominant advantage. All results are shown in Tables 10 and 11. On the Recall metric, MaterialAlphaSAM likewise attains the highest mean value, indicating that the proposed method is capable of effectively detecting the vast majority of target regions. Notably, while achieving the highest Recall, the proposed method simultaneously maintains a high level of Precision, signifying that the improvement in detection capability has not been accompanied by a substantial increase in false positives, thereby achieving a favorable balance between precision and recall.

**Table 10.** Overall quantitative performance comparison under a sample size of 10 (Mean ± Std across 5 folds).

| Method | IoU | F1-Score | Precision | Recall |
|---|---|---|---|---|
| MaterialAlphaSAM | 0.860 ± 0.013 | 0.924 ± 0.008 | 0.907 ± 0.010 | 0.943 ± 0.010 |
| SAM_Baseline | 0.620 ± 0.262 | 0.697 ± 0.276 | 0.848 ± 0.126 | 0.655 ± 0.281 |
| FCN | 0.609 ± 0.035 | 0.735 ± 0.025 | 0.752 ± 0.076 | 0.841 ± 0.077 |
| PSPNet | 0.550 ± 0.113 | 0.660 ± 0.122 | 0.855 ± 0.061 | 0.663 ± 0.149 |
| UNet | 0.471 ± 0.160 | 0.578 ± 0.190 | 0.795 ± 0.154 | 0.569 ± 0.185 |

**Table 11.** Overall quantitative performance comparison under a sample size of 50 (Mean ± Std across 5 folds).

| Method | IoU | F1-Score | Precision | Recall |
|---|---|---|---|---|
| MaterialAlphaSAM | 0.860 ± 0.012 | 0.924 ± 0.007 | 0.909 ± 0.006 | 0.943 ± 0.009 |
| SAM_Baseline | 0.620 ± 0.251 | 0.697 ± 0.260 | 0.848 ± 0.124 | 0.655 ± 0.266 |
| FCN | 0.609 ± 0.031 | 0.735 ± 0.023 | 0.752 ± 0.076 | 0.861 ± 0.056 |
| PSPNet | 0.471 ± 0.153 | 0.558 ± 0.150 | 0.795 ± 0.154 | 0.569 ± 0.170 |
| UNet | 0.550 ± 0.104 | 0.640 ± 0.113 | 0.854 ± 0.047 | 0.663 ± 0.133 |

In comparison, the SAM Baseline exhibits inferior performance across IoU, F1-Score, and Recall, with its primary bottleneck lying in severely deficient Recall. UNet, FCN, and

PSPNet outperform the SAM Baseline on certain individual metrics; however, their overall performance remains substantially inferior to that of MaterialAlphaSAM. SegNet records mean IoU and F1-Score values that are even lower than or comparable to those of the SAM Baseline, indicating that this architecture possesses inadequate adaptability for the present task.

#### 4.3.4. Per-Sample Performance Distribution Analysis

Figures 8 and 9 present the per-sample performance distributions of each method across all 5-fold test samples using violin plots, which provide finer-grained distributional characteristics beyond what summary statistics such as the mean can convey.

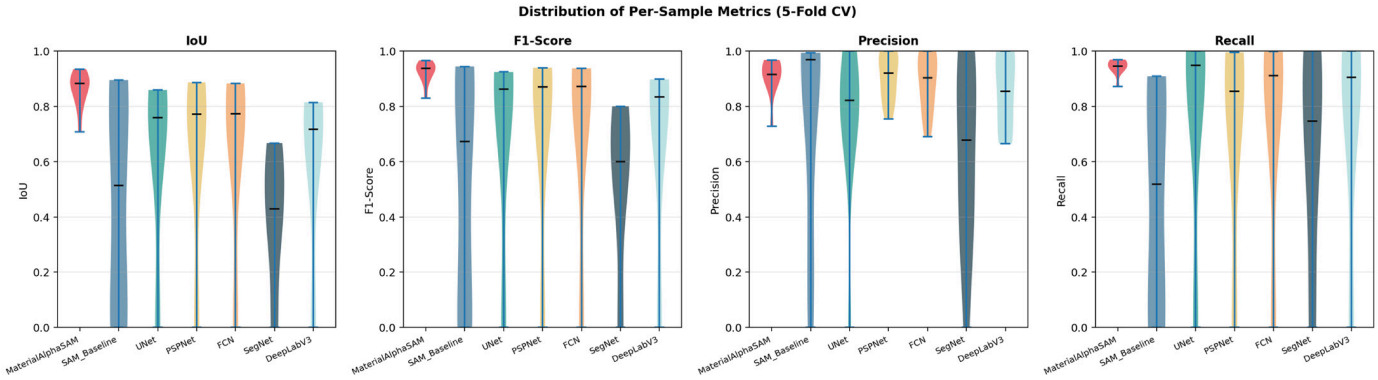



**Figure 8.** Per-sample performance distributions under a sample size of 10.

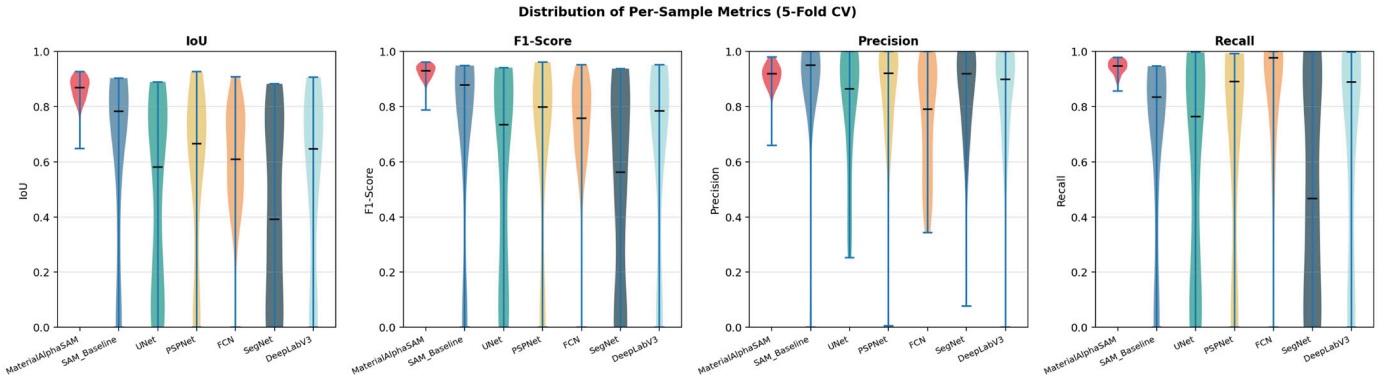



**Figure 9.** Per-sample performance distributions under a sample size of 50.

From the perspective of distributional morphology, MaterialAlphaSAM exhibits distributional characteristics that are markedly superior to those of all comparative methods across all four evaluation metrics. Specifically, the predominant mass of the proposed method's performance distribution is highly concentrated in the high-value region, with a median level noticeably above that of all comparison methods and a relatively narrow distributional width, indicating that high-quality segmentation results are achieved on the vast majority of individual samples. More importantly, MaterialAlphaSAM exhibits short distributional tails, with an extremely low proportion of extreme low-value samples, demonstrating that the method maintains stable segmentation capability across microscopic tissue images of varying morphologies and difficulty levels.

In comparison, the SAM Baseline exhibits an extremely dispersed distribution spanning nearly the entire value range on the IoU, F1-Score, and Recall metrics, with a comparatively low median level. This wide distributional spread reveals the severe instability of the original SAM at the individual sample level—while reasonable segmentation performance may be attained on a small number of relatively simple samples, significant performance

degradation occurs on a large proportion of samples, manifesting a pronounced bimodal polarization pattern. This micro-level distributional characteristic corroborates the macro-level catastrophic degradation observed in the per-fold performance analysis, thereby validating the vulnerability of the original SAM in industrial few-shot scenarios from a finer-grained perspective.

The distributional profiles of UNet, PSPNet, and FCN fall between those of MaterialAlphaSAM and the SAM Baseline. While their median levels are moderately improved relative to the SAM Baseline, a proportion of low-value tailing persists at the lower end of their distributions, indicating that the segmentation quality of these methods remains suboptimal on certain challenging samples. SegNet exhibits a generally low distributional profile, with considerable distributional width and low median levels across multiple metrics. DeepLabV3 presents a relatively wide distributional spread and moderate overall performance; however, its cross-sample consistency remains substantially inferior to that of the proposed method.

Taken together, MaterialAlphaSAM demonstrates two core advantages at the per-sample distributional level: first, a higher median performance level, reflecting a substantial improvement in overall segmentation quality; and second, a narrower distributional width and more compact tails, reflecting significantly enhanced cross-sample adaptability. The combination of these two advantages implies that the proposed method is capable of maintaining reliable segmentation quality across diverse categories of microscopic tissue images in practical deployment scenarios, rather than achieving favorable results only on specific sample types, thereby conferring stronger generalization capability and greater engineering practicality.

### *4.4. Comparison with SAM-Derived Adaptation Methods*

To further contextualize the contribution of the proposed framework within the broader landscape of recent SAM adaptation research, we extend the comparative evaluation to include two representative SAM-derived methods that are relevant to the present work: SAM-Adapter [18] and MedSAM [19].

Both methods are evaluated under a strictly unified experimental protocol consistent with all other baselines: identical fixed sample indices, identical 5-fold cross-validation partition scheme, and identical evaluation metric framework, across training sample scales of $n = 1000$. The quantitative results are presented in Table 12.

The experimental results reveal several noteworthy observations. First, MaterialAlphaSAM achieves superior performance over both SAM-Adapter and MedSAM across all four evaluation metrics under both sample scales, demonstrating that the explicit embedding of geometric morphological priors via the GCPP module, in conjunction with the lightweight domain adaptation of DAA, constitutes a more effective adaptation strategy for the present industrial microstructural segmentation task than general-purpose adapter insertion or large-scale domain fine-tuning alone. Second, the comparison between SAM-Adapter and MedSAM highlights the inherent limitations of domain-agnostic adaptation strategies when applied to highly specialized industrial image domains. Despite their demonstrated efficacy in medical and natural image contexts [20], neither method fully resolves the distributional gap between SAM's pre-training domain and the distinctive visual characteristics of high-magnification titanium alloy microstructural images, including weak phase boundaries, complex textural patterns, and heterogeneous morphological distributions. Furthermore, this gap persists even under large-scale domain fine-tuning when the source domain (medical imaging) remains structurally distant from the target domain (industrial metallography). Third, the parameter efficiency advantage of the proposed framework is further underscored by this comparison. MaterialAlphaSAM achieves

its performance superiority with a trainable parameter count substantially lower than that required by MedSAM's full fine-tuning scheme, confirming that the integration of domain-specific structural priors into the adaptation process constitutes a more principled and resource-efficient solution than scaling up the volume of fine-tuning parameters.

**Table 12.** Quantitative comparison including SAM-derived methods under a sample size of 1000 (Mean ± Std across 5 folds).

| Method | IoU | F1-Score | Precision | Recall |
|---|---|---|---|---|
| MaterialAlphaSAM | 0.864 ± 0.008 | 0.924 ± 0.021 | 0.910 ± 0.003 | 0.945 ± 0.021 |
| SAM-Adapter | 0.778 ± 0.028 | 0.865 ± 0.019 | 0.878 ± 0.025 | 0.858 ± 0.032 |
| MedSAM | 0.732 ± 0.045 | 0.835 ± 0.033 | 0.862 ± 0.021 | 0.815 ± 0.052 |

Collectively, these results provide quantitative evidence that the methodological contribution of the proposed framework extends meaningfully beyond what can be achieved through existing SAM-derived adaptation strategies, and lend further support to the broader argument that task-specific domain knowledge, when explicitly incorporated into the model adaptation process, serves as a more fundamental and efficient driver of performance improvement than general-purpose adaptation techniques in the context of specialized industrial visual inspection tasks.

## 5. Conclusions

This paper focuses on the material analysis task of fine segmentation for the primary $\alpha$ phase in high-magnification microstructure images and proposes a segmentation framework that integrates the Geometry-Constrained Prompt Prior and domain adaptation mechanism, namely, the MaterialAlphaSAM. Building on the strong global modeling capability of the SAM, this method systematically introduces a structure-aware prompt generation strategy and a lightweight domain adaptation module to tackle the inherent characteristics of microstructure images, including uneven scale distribution, highly complex morphologies and low phase boundary contrast, effectively bridging the semantic gap and feature distribution discrepancy between general visual models and fine-grained segmentation tasks in materials science.

The experimental results demonstrate that the proposed GCPP module can explicitly embed the geometric and morphological priors of microstructures into the prompt construction process, transforming prompt signals from uniform coverage to structure-aware guidance and significantly enhancing the spatial attention of the model to the core topological structures and blurred phase boundaries of target regions. The domain adaptation module effectively improves the training stability and segmentation robustness of the model under few-shot settings by alleviating the cross-domain feature distribution shift between natural images and microstructure images. Ablation studies further verify the functional complementarity of the GCPP and DAA modules; that is, the former provides stable and discriminative structured prompt constraints, whereas the latter enhances the domain consistency of feature representations. The synergistic effect enables the model to achieve the optimal overall performance at a moderate encoder unfreezing depth and yields significantly superior performance over traditional convolutional neural network methods and the SAM baseline strategy with regularized grid prompting key evaluation metrics, including the IoU and an F1-score.

Methodologically, the MaterialAlphaSAM not only significantly improves the accuracy and robustness of primary $\alpha$ phase microstructure segmentation, but also provides a generalizable technical paradigm for the deep adaptation of general visual foundation models in the field of materials science. This paradigm realizes high-fidelity parsing of com-

plex microstructural structures under limited annotation conditions by converting domain knowledge into structured prompt signals and combining them with a lightweight feature space adaptation mechanism. Nevertheless, there remains room for further optimization in terms of data diversity coverage, overfitting risk under deep fine-tuning conditions and inference efficiency.

**Author Contributions:** Conceptualization, K.L. and B.D.; methodology, K.L. and B.D.; formal analysis, Y.Z., W.L. (Wei Liu) and J.Z.; investigation, Y.Z., W.L. (Wei Liu) and J.Z.; resources, C.Y. and H.G.; writing—original draft preparation, K.L.; writing—review and editing, B.D. and D.T.; supervision, C.Y., D.T., H.G. and W.L. (Wenzhong Luo); project administration, W.L. (Wenzhong Luo). All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by the Key Research and Development Program of Shaanxi Province (No. 2025CY-YBXM-237, 2025CY-YBXM-542).

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

**Conflicts of Interest:** Author Ke Li, Bowen Deng, Yanru Zhao, Wei Liu, Chao Yang, Jing Zhu, Huixian Gao, and Wenzhong Luo were employed by the company Western Superconducting Technologies Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| SAM | Segment Anything Model |
| DAA | Domain-Adaptation Adapter |
| GCPP | Geometry-Constrained Prompt Prior |

## References

1. Tan, R.; Jin, S.; Wei, S.; Wang, J.; Zhao, X.; Wang, Z.; Liu, Q.; Sun, T. Evolution mechanism of microstructure and microhardness of Ti–6Al–4V alloy during ultrasonic elliptical vibration assisted ultra-precise cutting. *J. Mater. Res. Technol.* **2024**, *30*, 1641–1649.
2. Wang, K.X.; Zeng, W.D.; Shao, Y.T.; Zhao, Y.Q.; Zhou, Y.G. Quantification of microstructural features in titanium alloys based on stereology. *Rare Met. Mater. Eng.* **2009**, *38*, 398–403.
3. Campbell, A.; Murray, P.; Yakushina, E.; Marshall, S.; Ion, W. New methods for automatic quantification of microstructural features using digital image processing. *Mater. Des.* **2018**, *141*, 395–406. [CrossRef]
4. Zhao, P.; Wang, Y.; Jiang, B.; Wei, M.; Zhang, H.; Cheng, X. A new method for classifying and segmenting material microstructure based on machine learning. *Mater. Des.* **2023**, *227*, 111775. [CrossRef]
5. Biswas, M.; Pramanik, R.; Sen, S.; Sinitca, A.; Kaplun, D.; Sarkar, R. Microstructural segmentation using a union of attention guided U-Net models with different color transformed images. *Sci. Rep.* **2023**, *13*, 5737. [CrossRef] [PubMed]
6. Majumdar, S.; Sau, A.; Biswas, M.; Sarkar, R. Metallographic image segmentation using feature pyramid based recurrent residual U-Net. *Comput. Mater. Sci.* **2024**, *244*, 113199. [CrossRef]
7. Baskaran, A.; Kane, G.; Biggs, K.; Hull, R.; Lewis, D. Adaptive characterization of microstructure dataset using a two stage machine learning approach. *Comput. Mater. Sci.* **2020**, *177*, 109593. [CrossRef]
8. Fotos, G.; Campbell, A.; Murray, P.; Yakushina, E. Deep learning enhanced Watershed for microstructural analysis using a boundary class semantic segmentation. *J. Mater. Sci.* **2023**, *58*, 14390–14410. [CrossRef]
9. Thool, K.; Wi-Geol, S.; Dong-Ju, Y.; Choi, S.H. Deep learning-based automated phase segmentation of Zn-Al-Mg-Y microstructures. *Mater. Charact.* **2025**, *227*, 115303.
10. Gerçek, A.E.; Korten, T.; Chekhonin, P.; Hassan, M.; Steinbach, P. Data-efficient U-Net for segmentation of carbide microstructures in SEM images of steel alloys. *arXiv* **2025**, arXiv:2511.11485.
11. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 3992–4003.

12. Fan, Q.; Tao, X.; Ke, L.; Ye, M.; Zhang, Y.; Wan, P.; Wang, Z.M.; Tai, Y.W.; Tang, C.K. Stable segment anything model. *arXiv* **2023**, arXiv:2311.15776.
13. Zhang, M.; Dai, X.; Sun, Y.; Wang, J.; Yao, Y.; Gong, X.; Cong, F.; Wang, F.; Lv, Y. Hierarchical self-prompting SAM: A prompt-free medical image segmentation framework. *arXiv* **2025**, arXiv:2506.02854.
14. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; pp. 234–241.
16. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [PubMed]
17. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
18. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
19. Chen, T.; Zhu, L.; Deng, C.; Cao, R.; Wang, Y.; Zhang, S.; Li, Z.; Sun, L.; Zang, Y.; Mao, P. SAM-Adapter: Adapting Segment Anything in Underperformed Scenes. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Paris, France, 2–6 October 2023; pp. 3367–3375.
20. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* **2024**, *15*, 654. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.