Efficient Domain Knowledge Injection for Bridging the Gap Between Generalized Large Vision Models and Specialized Fabric Defect Tasks

Zhewei Chen, Wai Keung Wong, Zuofeng Zhong, Jinpiao Liao & Ying Qu

View supplementary material

Published online: 26 Sep 2024.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Efficient Domain Knowledge Injection for Bridging the Gap Between Generalized Large Vision Models and Specialized Fabric Defect Tasks

Zhewei Chen[a], Wai Keung Wong[a,b], Zuofeng Zhong[b], Jinpiao Liao[a], and Ying Qu[a]

[a]School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong SAR, China; [b]Laboratory for Artificial Intelligence in Design, Hong Kong SAR, China

ABSTRACT

The scarcity of high-quality annotated data poses a significant challenge to the application of deep learning in fabric defect tasks, limiting the generalization and segmentation performance of existing models and impeding their capability to address the complexity of various fabric types and defects. To overcome these obstacles, this study introduces an innovative method to infuse specialized knowledge of fabric defects into the Segment Anything Model (SAM), a large-scale visual model. By introducing and training a unique set of fabric defect-related parameters, this approach seamlessly integrates domain-specific knowledge into SAM without the need for extensive modifications to the preexisting model parameters. The revamped SAM model leverages generalized image understanding learned from large-scale natural image datasets while incorporating fabric defect-specific knowledge, ensuring its proficiency in fabric defect segmentation tasks. The experimental results reveal a significant improvement in the model's segmentation performance, attributable to this novel amalgamation of generic and fabric-specific knowledge. When benchmarking against popular existing segmentation models across three datasets, our proposed model demonstrates a substantial leap in performance. Its impressive results in cross-dataset comparisons and few-shot learning experiments further demonstrate its potential for practical applications in textile quality control.

摘要

高质量带注释数据的稀缺性对深度学习在织物缺陷任务中的应用提出了重大挑战，限制了现有模型的泛化和分割性能，并阻碍了它们解决各种织物类型和缺陷复杂性的能力. 为了克服这些障碍，本研究引入了一种创新方法，将织物缺陷的专业知识注入到大规模视觉模型分段任意模型（SAM）中. 通过引入和训练一组独特的织物缺陷相关参数，这种方法将特定领域的知识无缝集成到SAM中，而不需要对预先存在的模型参数进行大量修改. 改进后的SAM模型利用从大规模自然图像数据集中学习到的广义图像理解，同时结合织物缺陷特定知识，确保其在织物缺陷分割任务中的熟练程度. 实验结果表明，由于这种通用知识和织物特定知识的新颖融合，模型的分割性能得到了显著提高. 当在三个数据集上对流行的现有分割模型进行基准测试时，我们提出的模型在性能上实现了巨大的飞跃. 它在跨数据集比较和少数镜头学习实验中的令人印象深刻的结果进一步证明了它在纺织品质量控制中的实际应用潜力.

Introduction

FABRIC defect segmentation plays a crucial role in the quality control of textiles, as it not only detects defects but also precisely outlines their shapes and sizes. This level of detail enables more targeted and efficient remediation processes, ultimately ensuring higher quality control and product integrity in the textile industry (Boluki and Mohanna 2021). Fabric defects, such as holes, stains, and broken yarns, are the main factors affecting fabric quality. Therefore, effectively and accurately identifying and segmenting defects in fabrics is of great significance for ensuring the quality of textiles, improving production efficiency, and reducing production costs (C. Li et al. 2021). Especially in modern textile industry, with the expansion of production scale and the improvement of efficiency, relying on manual identification and segmentation of fabric defects may not satisfy production needs (S. Zhao et al. 2020). Thus, it is crucial to develop a fully automatic technology for accurate and efficient fabric defect segmentation (Ngan, Pang, and Yung 2011).

Despite the noteworthy successes of deep learning methods in various fabric defect datasets (Arora and Hanmandlu 2022; T. Chen et al. 2015; Y. Li, Zhao, and Pan 2016; Peng et al. 2020; Tian and Li 2019; Zhu et al. 2020), performance of fabric defect segmentation still relies on the scale of training data and the parameters of the model (Figure 1a). Apparently, the scarcity of high-quality annotated data, compounded by the wide range of fabric defects and types (Huang, Jing, and Wang 2021), limits the application of deep learning in this field. Given the extensive diversity of fabric defects and types, creating a sufficient annotated dataset encompassing all possible combinations is practically unattainable (Kahraman and Durmuşoğlu 2023). This insufficient data thereby hinders the progress of models designed for fabric segmentation tasks. Although, models with large parameters can capture more complex image features, they are prone to overfitting on small-scale training data (Ying 2019), resulting in a decrease in segmentation performance in practical applications. By contrast, small-parameter models can achieve
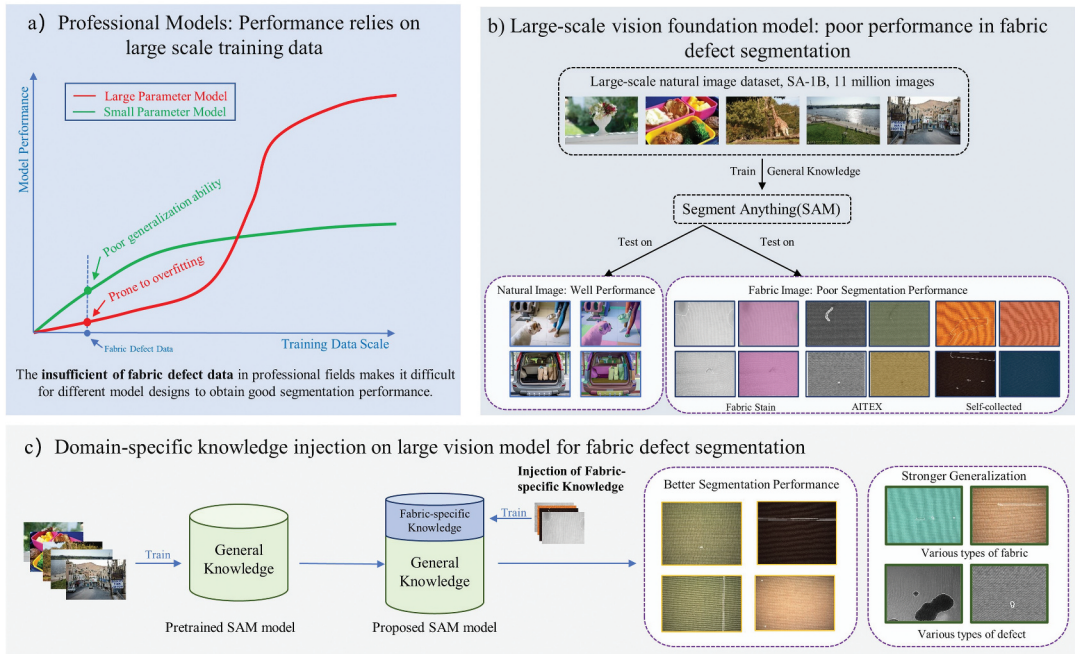
Figure 1. Motivation for injecting fabric-specific knowledge in segment anything model (SAM). a) the scarcity of datasets in the fabric defect domain restricts the performance of specialized models trained solely on fabric defect data. b) large-scale vision models, trained on vast natural image datasets, lack specialized knowledge of fabric defects, resulting in subpar performance in fabric defect segmentation tasks. c) this paper proposes the injection of specialized fabric defect knowledge into large-scale vision models to enhance their performance and generalization abilities in fabric defect segmentation tasks.

good results on small-scale training data, but they often struggle to obtain well generalization (Canatar, Bordelon, and Pehlevan 2021), especially when tackling diverse fabric types and defect categories.

Recently popular Segment Anything Model (SAM) (Kirillov et al. 2023), a large-scale vision foundation model, presents a potential solution to overcome the above challenges. It leverages training on billions of natural images to accumulate a rich reservoir of generalized knowledge (Han et al. 2021). This wide-ranging understanding enhances its ability to decode complex image structures and patterns, boosting its generalization capacity. Excelling across diverse image tasks, SAM has earned significant acclaim and wide applications (R. Deng et al. 2023; Ji et al. 2023; Y. Jing, Wang, and Tao 2023; Yu et al. 2023). Despite the abundant general knowledge inherent within the SAM, its direct application to fabric defect segmentation frequently results in suboptimal performance (Figure 1b). This deficiency is principally attributed to the model's insufficiency in specialized knowledge of fabric defects. The general knowledge gained from training on diverse natural images does not effectively translate into identifying and segmenting fabric defects. Therefore, the main challenge of transferring the general model to fabric defect segmentation task is how to enhance it with fabric defect-specific knowledge while preserving its extensive general knowledge.

To address the aforementioned challenges, a specialized approach is proposed to enhance the SAM with domain-specific knowledge pertaining to fabric defects (Figure 1c). In this approach, an additional set of parameters associated with fabric defects is introduced into SAM, utilizing a distinctive training strategy. This strategy, utilizing specific constraints during the training phase, steers these newly introduced parameters toward learning the unique characteristics of fabric defects. Concurrently, these parameters exert subtle influence over the original parameters of SAM, directing them toward an understanding of fabric defects, while conserving their foundational comprehension of visual knowledge. SAM thus functions as a hybrid model, amalgamating the merits of a large-scale visual model and a task-specific small model. It leverages SAM's extensive generic visual understanding sourced from a variety of natural images, and supplements it with the precise, fabric defect-specific knowledge delivered by the additional parameters. The proposed approach showcases its efficacy in scenarios with limited training data, enhancing the model's robust segmentation performance and generalization capability in the face of data constraints. The contributions of this study are as follows:

1) To the best of our knowledge, we are the first to successfully employ a large-scale visual model to the field of fabric defect segmentation. By fully considering the characteristics of the domain, such as the rich texture of textiles, the diversity of defect types, and the scarcity of samples, our method can achieve more accurate and effective detection of fabric defects. This not only enhances the precision and efficiency of fabric defect detection but also provides new insights for research in this field.

2) We introduced a novel method that incorporates specialized knowledge of fabric defects into the SAM, maintaining its universal visual comprehension while infusing domain-specific insights. Proposed method capitalizes on the general knowledge base of large visual model to mitigate issues arising from defect data scarcity.

3) The proposed model has been extensively benchmarked against popular segmentation models across three fabric defect datasets. Supported by a comprehensive set of experimental results, the proposed model has demonstrated outstanding defect segmentation capabilities and impressive generalization ability, thus establishing a significant advancement in fabric quality control.

This paper is structured as follows: Section 2 reviews related literature. Section 3 elucidates the proposed approach to infuse specialized knowledge into the SAM model. Section 4 describes the experimental design. Sections 5, 6, and 7 present the experimental results, discuss their implications, and summarize the study, respectively.

Related works

Fabric defect segmentation

Deep learning has significantly surpassed traditional methods in the task of fabric defect segmentation, with most current methods being based on this approach. These studies can be categorized into two

types based on the availability of pixel-wise labeled data: supervised learning and weakly supervised learning. The first category involves supervised learning methods that require pixel-wise labeled data for model training. The second category includes methods that either do not use pixel-wise labeled data or only employ defect-free samples for training. This can be considered as a type of weakly supervised learning method.

In the supervised learning, due to the use of pixel-wise training data, many studies have achieved remarkable segmentation performance on corresponding datasets. For example, Cheng et al. (2023) proposed a fabric defect detection method based on Separate Convolutional UNet (SCUNet) and achieved 98.01% accuracy on AITEX dataset. To achieve real time segmentation, Huang et al. (2021) proposed a highly efficient convolutional neural network for fabric defect detection, which achieved high-accuracy defect localization with as few as 50 defect samples and real-time detection speeds, outperforming eight state-of-the-art methods in terms of accuracy and robustness. To further improve performance, Liu et al. (2019) proposed a deep saliency model with an attention mechanism for fabric defect detection, successfully enhancing detection performance and surpassing state-of-the-art methods in defect localization without significantly increasing computational demands. Similar studies based on deep learning and supervised learning for fabric defect detection include (J. Jing et al. 2022; Kopaczka et al. 2019; Y. Li, Zhao, and Pan 2016; Ouyang et al. 2019; Rong-Qiang et al. 2021; Shao et al. 2022; Sun et al. 2019). However, supervised learning heavily relies on large-scale and high-quality labeled data. Collecting and annotating fabric defect images can be challenging, making it impractical to rely solely on supervised learning for fabric defect segmentation.

Some researchers have tried to applied weakly supervised learning methods to resolve the issue of pixel-wise annotation acquisition by utilizing more defect-free image information and other types of annotation information for defect segmentation. Koulali and Eskil (2021) proposed a method that only need one defect-free sample for training and demonstrated competitive performance on the Patterned Fabrics benchmark dataset. J. Liu et al. (2019) proposed a GAN-based framework for fabric defect detection that customizes a deep semantic segmentation network for various defect types and utilizes a multistage GAN to generate reasonable defects on new defect-free samples. Their approach enables continuous dataset updates and contributes to the fine-tuning of the network to enhance defect detection performance under varying conditions. However, these methods often struggle with lower accuracy in defect segmentation due to the difficulty in obtaining precise locations and shapes of defects compared to fully supervised methods. Model training can also be challenging, as weak supervision relies on limited pixel-level annotation data, requiring more iterations and complex optimization algorithms. Performance across different datasets can be unstable, sometimes matching or surpassing supervised methods, while underperforming in other cases (B. Li et al. 2022; Niu et al. 2022).

Applications build upon segment anything model (SAM)

SAM has been applied in various domains, including medical image segmentation, natural scene segmentation, and video object segmentation. For instance, R. Deng et al. (2023) evaluated the SAM for digital pathology tasks and demonstrates that SAM performs well for large connected objects but struggles with dense instance objects. The authors discussed the limitations of SAM in the context of pathological images and proposed future directions for its application. Similarly, Tang, Xiao, and Li (2023) investigated the applicability of the SAM for the task of camouflaged object detection (COD). They evaluated SAM on a COD benchmark using two metrics and compared its performance against 22 state-of-the-art COD methods. The results revealed that SAM exhibits limited performance in the COD task, prompting further research to enhance its generalization to specific scenes. Additionally, in studies (Ma et al. 2023; Wu et al. 2023; Zhang et al. 2023), the authors attempted to build foundation model for medical images based on SAM and demonstrated its excellent performance on downstream medical tasks. Similar studies (Cen et al. 2023; Larsen et al. 2023; Mo and Tian 2023; Zhou et al. 2023) also highlight that while SAM demonstrates impressive performance in the domain of natural images,

its segmentation effectiveness diminishes when applied to specific fields such as RGBD images, electron microscopy images and remote sensing images.

To further emphasize the significance of this research, this study utilizes the presence of the rich and generic knowledge in pretrained SAM to compensate for the limited segmentation performance and generalization caused by the insufficient training data in fabric defect segmentation. By integrating the general knowledge from SAM, the generalization performance of the fabric defect segmentation model is significantly enhanced. Moreover, the learning of specialized fabric defect knowledge ensures the model's ability to accurately recognize fabric defects. Compared to traditional supervised and unsupervised methods, SAM offers several advantages: it requires less annotated data while leveraging extensive pre-trained knowledge, demonstrates stronger generalization across various image domains, provides simplicity and efficiency through prompt-based segmentation, and allows for domain adaptability when combined with specific fabric defect knowledge. These benefits address key limitations of conventional approaches in fabric defect detection, such as data scarcity, inconsistent performance across datasets, and complex optimization requirements. To the best of our knowledge, this research is the first to study the application of large-scale pretrained models to the task of fabric defect segmentation, potentially opening new avenues for improved defect detection in the textile industry.

Methods

Architecture of our proposed method is built upon the pretrained SAM (Figure 2). Additional trainable parameters are introduced to the self-attention mechanism in the transformer module of the SAM, aiming to inject specialized knowledge related to fabric defects. In the following sections, the structure of the SAM, which serves as a large-scale model, will be discussed. Subsequently, the process of introducing the fabric defect knowledge-related embeddings, as introduced in this paper, will be elaborated. Finally, the optimization process for the entire model will be outlined.
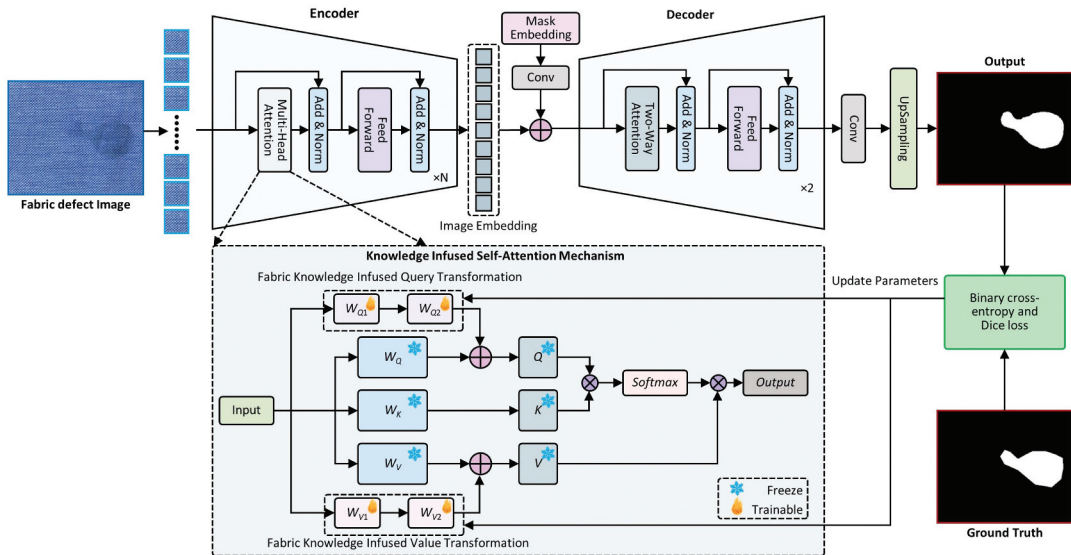
Figure 2. Architecture of the proposed model, with specific modifications made to the self-attention mechanism within the encoder and decoder's transformer modules. Additional matrices, parallel to W_Q and W_K, representing the injection of specialized knowledge, have been incorporated into the model. During the training process, only the parameters within these specialized knowledge matrices in the encoder are trainable while other parameters remain frozen. Additionally, all parameters within the decoder are updated.

Segment anything model (SAM)

SAM model is mainly composed of three core components: an Image Encoder, a Prompt Encoder, and a Mask Decoder.

The Image Encoder adopts the classical Vision Transformer architecture (Dosovitskiy et al. 2020). The input image is encoded into image embeddings at a scale of 16×16. During the encoding process, the model introduces two types of attention mechanisms: local and global attention. In local attention, each patch interacts only with other patches within a set window size. Global attention, on the other hand, allows each patch to interact with all other patches. Within the SAM model, the Image Encoder module possesses the largest quantity of parameters. For different versions of SAM (e.g., SAM-B, SAM-L, SAM-H), the model is differentiated by adjusting the embedding dimension and encoder depth; the larger the model scale, the higher the embedding dimension, and the greater the encoder depth.

The role of the Prompt Encoder is to encode image prompts (such as points, boxes, or masks) into a format that can be processed by the decoder. In order to retain the original spatial information, this encoder also adds position encoding. In our experiments, aside from the default dense mask prompts, we did not use other types of prompts. This means that the models in this study do not support interactive segmentation like SAM.

The Mask Decoder is a specially designed Transformer encoder structure that includes a unique dynamic mask generation submodule. In the original SAM model, each encoder block innovatively uses a bidirectional cross-attention mechanism, rather than the traditional self-attention mechanism found in typical Transformers. This mechanism allows for bidirectional interaction from prompts to image embeddings and from image embeddings to prompts, effectively learning the relationship between prompts and image embeddings within the model. After passing through two such encoder blocks, the SAM model upsamples the image embeddings to increase resolution. Then, through a Multi-Layer Perceptron (MLP), the model maps the output tokens to a dynamic linear classifier, responsible for predicting the mask of the target image. It should be noted that the generated mask is only 1/4 the size of the original image, so in practical application, it is upsampled to the same size as the original image. In this study, we used the Mask Decoder from the original SAM model without any modifications to its structure. Figure 2 shows the complete SAM model structure we used, including this innovative Mask Decoder module.

Fabric defect knowledge-related parameters

Fabric defect knowledge-related parameters are injected to enhance the performance on fabric defect segmentation and preserve the general knowledge of SAM. It can prevent catastrophic forgetting during the learning of fabric defect-specific knowledge. To maintain the architecture of original SAM, we only modified the self-attention mechanism in the transformer module to incorporate the general knowledge and fabric defect-specific knowledge.

In transformer module, the Q, K, and V weight matrices carry the general semantic information of the large-scale model. They play a crucial role in the multi-head attention mechanism, where they compute the matching degree between queries (Q) and keys (K) to determine attention weights. These weights are then applied to the value (V) matrix, generating weighted semantic representations that capture long-range dependencies in the input sequence. These weights can be viewed as a set of encoding functions learned by the model from the input sequence, mapping the inputs into a semantic space and encompassing the model's learning and understanding of general semantic information across different tasks and domains. Consequently, large-scale transformer models exhibit strong generalization and transfer learning capabilities, delivering outstanding performance across various tasks and domains.

In the original SAM model, the generation process of the Q, K, and V keys in the self-attention mechanism can be expressed using the following equations:

$$Q = X \cdot W_Q \tag{1}$$

$$K = X \cdot W_K \tag{2}$$

$$V = X \cdot W_V \tag{3}$$

Here, W_Q, W_K, and W_V are weight matrices that are multiplied with the input features X to generate the corresponding Q, K, and V keys. This process leverages the properties of matrix multiplication to map the input features to the space of Q, K, and V keys.

In this study, the low-rank adaptation technique (Hu et al. 2021) is employed to modify the generation of Q, K, and V keys in the model. This technique involved adding a bypass to the original weight matrices to achieve low-rank approximation. The schematic diagram of knowledge infused self-attention mechanism in Figure 2 illustrates this process. The generation of Q, K, and V keys can be formulated as the combination of the original weight matrix and additional linear layers W_{Q1}, W_{Q2}, W_{V1}, and W_{V2} obtained from the bypass. It can be formulated as:

$$Q = X \cdot W_{Q1} \cdot W_{Q2} + X \cdot W_Q \tag{4}$$

$$K = X \cdot W_K \tag{5}$$

$$V = X \cdot W_{V1} \cdot W_{V2} + X \cdot W_V \tag{6}$$

In the above formulas, X represents the input features. The output of W_{Q1} and W_{Q2} actually forms a new weight matrix that represents the newly learned knowledge specific to the fabric defect detection task. In this context, W_{Q1} and W_{Q2} can be seen as components of the weight matrix, which are trained to capture specific features relevant to fabric defect detection. In this setup, the original weight matrix W_Q retains the general knowledge learned from pre-trained SAM, while $W_{Q1} \cdot W_{Q2}$ provides task-specific knowledge for fabric defect detection. In this case, the dimensions of W_{Q1} and W_{Q2} can be represented as $W_{Q1} \in \mathbb{R}^{in \times r}$ and $W_{Q2} \in \mathbb{R}^{r \times out}$, respectively. Here, in represents the dimensionality of the input features, out represents the dimensionality of the output features, and it should match the dimensionality of W. Where $R = W_{Q1} \cdot W_{Q2}$ is a low-rank matrix learned during the training process, with dimensions $\mathbb{R}^{r \times out}$. The specific values of each R vary according to input features and task requirements. This allows the model to dynamically adapt to different fabric defect detection tasks. The dimensions of W_{Q1} and W_{Q2} are $\mathbb{R}^{in \times r}$ and $\mathbb{R}^{r \times out}$ respectively, where in is the dimension of the input features, and out is the dimension of the output features, which should match the dimensions of the original weight matrix W. In this setup, r is a preset hyperparameter representing the dimension of the low-rank adaptation. This is a hyperparameter, and in this study, we set $r = 4$. However, the optimal value of r may vary depending on the specific task and dataset.

After obtaining the new Q, K, and V matrices, the self-attention mechanism remains consistent with the original SAM and can be expressed by the following equation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

Here, Q, K, and V represent the query matrix, key matrix, and value matrix, respectively. By calculating the similarity between the queries and keys, normalizing the similarity scores using the Softmax function, and applying the resulting attention weights to the value matrix, a weighted semantic representation is generated.

Training strategy

The fabric defect segmentation problem can be formulated as a pixel-wise binary classification problem. Given an input fabric image I with a resolution of $H \times W$, the model generates a segmentation map S of the same resolution. Each pixel in S is classified as either defective or non-defective. Mathematically, the function that our model needs to learn can be described as follows:

$$F : I \in \mathbb{R}^{H \times W \times 3} \rightarrow S \in \{0, 1\}^{H \times W} \tag{8}$$

Here, F is the function implemented by our model, 3 is the number of color channels in the input image, and $\{0, 1\}$ represents the binary label for each pixel, where 0 denotes non-defective and 1 denotes defective.

During the training process, we modified each self-attention layer in both the encoder to the structure illustrated in Figure 2, while the self-attention layers in the decoder remained unchanged. In the encoder, except for our newly added layers, all parameters are fixed. As for the decoder, all parameters are updated.

A composite loss function that combines the Dice loss and Binary Cross-Entropy (BCE) loss is used to optimize the model, effectively balancing the class and spatial information. The overall loss function L can be represented as:

$$L = \alpha L_{BCE} + (1 - \alpha) L_{Dice} \tag{9}$$

Here, L_{BCE} refers to the BCE loss, calculated as follows:

$$L_{BCE} = -\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [y_{ij} \log(p_{ij}) + (1 - y_{ij}) \log(1 - p_{ij})] \tag{10}$$

Where y_{ij} denotes the label of the pixel at location (i, j), and p_{ij} is the predicted probability that the pixel at location (i, j) is defective.

L_{Dice} refers to the Dice loss, computed as follows:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^{H} \sum_{j=1}^{W} y_{ij} p_{ij} + \varepsilon}{\sum_{i=1}^{H} \sum_{j=1}^{W} y_{ij}^2 + \sum_{i=1}^{H} \sum_{j=1}^{W} p_{ij}^2 + \varepsilon} \tag{11}$$

Where ε indicates a small number added to prevent division by zero, and α denotes a weight parameter controlling the relative importance of the two loss components.

Experiments

Experimental data

In this study, three fabric defect detection datasets: the Fabric Stain dataset (Pathirana 2020), the AITEX (Silvestre-Blanes et al. 2019) dataset, and a self-collected dataset are utilized to evaluate our proposed method. The Fabric Stain and AITEX datasets are publicly available, while the self-collected dataset was procured from a weaving factory. All images used in this study contain defects, and for the AITEX and Fabric Stain datasets, we specifically selected only the images with defects for our experiments.

The Fabric Stain dataset originally provides bounding box annotations for defects. Subsequently, pixel-level annotations were generated based on these bounding boxes by a expert employing the LabelMe tool to manually delineate the precise contours within each bounding box, thereby providing detailed pixel-level defect annotations. Both the AITEX and self-collected datasets include pixel-level annotations.

The datasets are summarized as follows: The Fabric Stain dataset comprises 394 defective images and corresponding labels, the AITEX dataset contains 185 defective images and labels, and the self-collected dataset consists of 1515 defective images and labels. Each dataset was then partitioned into distinct subsets: 60% of the images for training, 20% for validation, and the remaining 20% for testing.

Evaluation metrics

To assess fabric defect segmentation, four widely-used metrics, recall, precision, Dice coefficient (Dice), and Intersection over Union (IoU), are used:

Recall measures the proportion of correctly identified positive samples:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

Precision assesses the proportion of correctly identified positive predictions:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

Dice Coefficient gauges the overlap between the predicted and actual segmentation:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{14}$$

Intersection over Union (IoU) evaluates the ratio of the intersection to the union of predicted and actual areas:

$$IoU = \frac{TP}{TP + FP + FN} \tag{15}$$

Where TP is True Positive, FP is False Positive, and FN is False Negative.

Training process

In all experiments, the inputs to all models are unified to 512×512 using the letterbox method to facilitate cross-validation across different datasets. If the original image size is less than 512×512, no magnification is performed, and padding is added instead.

The optimization process of all models, including our proposed SAM, employs the AdamW (Loshchilov and Hutter 2017) optimizer with momentum correction to update the model parameters, where the learning rate is set to 1e-3 and the beta values are set to 0.937 and 0.999. The training procedure uses a mini-batch of 8 samples and is executed on a GPU with 24GB of memory. If a single GPU cannot handle a batch of 8 images at once, we use a gradient accumulation approach for updates. A cosine annealing warm restarts learning rate scheduler is also employed in this study. It has a period of 10, a period multiplier of 100, and a minimum learning rate set to one percent of the initial learning rate. The model is trained for 1000 epochs and utilizes an early stopping strategy, saving checkpoints at the epoch where the model achieves the best validation Dice score. All non-SAM models were implemented with a ResNet101 (He et al. 2016) backbone and initialized with weights pre-trained on the ImageNet dataset (J. Deng et al. 2009).

During the training process, data augmentation techniques are applied to prevent overfitting. Specifically, the augmentation pipeline includes horizontal and vertical flipping of images, each with a probability of 0.5. Additionally, a random 90-degree rotation is applied with a probability of 0.5. Moreover, random adjustments of brightness and contrast are utilized, each within a range of 0.2 and with a likelihood of 0.5.

Cross-dataset validation

To evaluate the generalization of the model, each model was trained on one dataset and subsequently tested on the other two datasets. For example, when the privately collected dataset was used for training, the entire Fabric Stain and AITEX datasets were used for testing. This procedure was then repeated for the other datasets, providing a robust measure of the model's generalization capability.

During the validation process, the original hyperparameters and training strategies used during the training phase were strictly maintained. This approach was crucial in ensuring that the model's performance on different datasets could be attributed to its generalization ability, rather than over-fitting or specific adjustments tailored to a particular dataset.

Few-shot experiments

Few-shot experiments aim to evaluate the performance of the SAM model in limited-sample learning scenarios. Specifically, the three variants of the SAM (SAM-B, SAM-L, and SAM-H) were tested in three different scenarios (10 samples, 50 samples, and 100 samples) using our privately collected dataset.

In each scenario, a random selection of 10, 50, or 100 samples was made from the training set of the privately collected dataset for model training. Subsequently, the model was tested on the original test set. The training process in each scenario remained consistent with the process used when utilizing the full image set.

Results

We conducted a comprehensive comparative analysis of the segmentation performance of the proposed models (SAM-B, SAM-L, SAM-H) against several other prominent models in the field, such as Unet (Ronneberger, Fischer, and Brox 2015), Unet++ (Zhou et al. 2018), Manet (Fan et al. 2020), Linknet (Chaurasia and Culurciello 2017), FPN (Kirillov et al. 2017), PSPNet (H. Zhao et al. 2017), PAN (H. Li et al. 2018), DeepLabV3 (L.-C. Chen et al. 2017), DeepLabV3+ (L.-C. Chen et al. 2018), SCUnet (Cheng et al. 2023), Mobile-Deeplab (Bai and Jing 2023) and Swin-Unet (Xu et al. 2023). The models were tested on three different fabric defect detection datasets: Stain, AITEX, and a self-collected dataset.

Model performance across internal datasets

The results in Table 1 along the diagonal illustrate that our proposed model achieved outstanding performance on the internal datasets. For the Stain Dataset, the proposed SAM-H model significantly outperforms all other models across all performance metrics. It achieves substantial improvements in recall, precision, F1 score, and IoU compared to the next best-performing models. Specifically, the SAM-H model surpasses other models by 3.6% in recall, 7% in precision, 5.8% in F1 score, and 8% in IoU. On the AITEX dataset, the proposed SAM models continue to exhibit the highest segmentation performance. Notably, the SAM-L model yields the best results in terms of recall and IoU, while the SAM-B model achieves the best precision and F1 score. In terms of segmentation metrics, the proposed SAM models outperform the non-SAM models by 4.2%, 0.6%, 4.2%, and 3.8%, respectively. For the self-collected dataset, the proposed SAM models again obtain the best segmentation results. In particular, the SAM-H model demonstrates superior performance in terms of precision and F1 score, while the SAM-L model achieved the best recall and IoU. The proposed SAM models exhibit improved segmentation performance compared to the non-SAM models, with enhancements of 2.6% in recall, 6.3% in precision, 5.1% in IoU, and 7.2% in F1 score.

The visual results depicted in Figure 3 demonstrate that, when trained and tested on the self-collected dataset, our proposed method exhibits significantly better segmentation performance compared to other models, particularly for unclear defects (as shown in the second row of Figure 3) and subtle fabric defects (illustrated in the third row of Figure 3). From the results presented, it is evident that the proposed SAM models considerably outperform the other methods across all metrics and datasets. This underpins the precision and effectiveness of the proposed approach in fabric defect segmentation.

Table 1. Comparison of segmentation performance of various methods on different datasets.

Train on	Method	Test on											
		Stain				AITEX				Self-collected			
		recall	precision	F1	iou	recall	precision	F1	iou	recall	precision	F1	iou
Stain	Unet	0.8212	0.7328	0.7452	0.6452	0.5386	0.1278	0.1144	0.0812	0.4873	0.0759	0.0641	0.0450
	Unet++	0.8431	0.7640	0.7748	0.6849	0.5765	0.3908	0.1539	0.1202	**0.5761**	0.2590	0.1833	0.1280
	Manet	0.8496	0.7694	0.7716	0.6783	0.5476	0.5156	0.0765	0.0758	0.5403	**0.4087**	0.0953	0.0883
	Linknet	0.8318	0.7609	0.7769	0.6743	0.5777	0.2296	0.1922	0.1201	0.5331	0.1811	0.1613	0.0911
	FPN	0.8502	0.7606	0.7646	0.6871	0.5896	0.3305	0.1348	0.1192	0.5688	0.2407	0.1927	0.1378
	PSPNet	0.8210	0.7247	0.7333	0.6357	0.5775	0.2613	0.0797	0.0753	0.4747	0.0141	0.0088	0.0061
	PAN	0.8538	0.7449	0.7263	0.6510	0.5555	0.3151	0.0906	0.0846	0.4760	0.0255	0.0247	0.0150
	DeepLabV3	0.8630	0.7656	0.7598	0.6884	**0.6011**	0.2167	0.1071	0.0909	0.5146	0.1307	0.1289	0.0785
	DeepLabV3+	0.8254	0.7467	0.7628	0.6646	0.5929	0.4120	0.1064	0.1000	0.5565	0.2248	0.1295	0.0921
	SCUnet	0.8308	0.7425	0.7485	0.6495	0.0109	0.0109	0.0207	0.0109	0.0580	0.0580	0.0847	0.0580
	Mobile-Deeplab	0.7828	0.7446	0.7350	0.6279	0.6850	0.0655	0.0843	0.0517	0.1425	0.1017	0.0922	0.0759
	Swin-Unet	0.5838	0.6638	0.5639	0.4424	0.7400	0.0500	0.0758	0.0456	0.2123	0.0725	0.0532	0.0327
	Proposed SAM-B	0.8678	0.7945	0.7978	0.7157	0.5609	0.1864	0.1859	0.1414	0.5418	0.1863	0.1903	0.1401
	Proposed SAM-L	0.8911	0.8215	0.8193	0.7536	0.5885	0.2294	0.2201	0.1748	0.5547	0.2191	0.2254	0.1673
	Proposed SAM-H	**0.8992**	**0.8356**	**0.8327**	**0.7673**	0.5358	**0.5225**	**0.2269**	**0.1869**	0.5588	0.2322	**0.2391**	**0.1777**
AITEX	Unet	0.4659	0.0669	0.0690	0.0081	0.6625	0.4025	0.4090	0.3162	0.4954	0.1262	0.1108	0.0381
	Unet++	0.4976	0.1980	0.2100	0.0790	0.6978	0.4940	0.4709	0.3604	0.5007	0.2175	0.1094	0.0507
	Manet	0.5040	0.1958	0.2084	0.0900	0.6935	0.4695	0.4558	0.3354	0.4762	0.0604	0.0556	0.0211
	Linknet	0.4903	0.1827	0.1939	0.0622	0.6645	0.4159	0.4123	0.3029	0.4867	0.1300	0.1332	0.0382
	FPN	0.4801	0.0885	0.0904	0.0327	0.6804	0.4459	0.4479	0.3466	0.4980	0.1670	0.1545	0.0521
	PSPNet	0.4719	0.0350	0.0377	0.0197	0.6197	0.3025	0.3069	0.2413	0.4760	0.0464	0.0486	0.0168
	PAN	0.4821	0.1204	0.1155	0.0237	0.6692	0.4591	0.4786	0.3482	0.4745	0.0439	0.0453	0.0149
	DeepLabV3	0.4713	0.0450	0.0420	0.0138	0.5955	0.2985	0.3088	0.2111	0.5097	0.1544	0.1558	0.0728
	DeepLabV3+	0.5392	0.2236	0.2212	0.1320	0.6515	0.4244	0.4173	0.2767	0.5393	0.1774	0.1437	0.0712
	SCUnet	0.0533	0.0533	0.0952	0.0533	0.6949	0.4636	0.4908	0.3859	0.0579	0.0579	0.0846	0.0579
	Mobile-Deeplab	0.0075	0.1226	0.0131	0.0072	0.3378	0.3223	0.3074	0.2540	0.0107	0.0449	0.0104	0.0067
	Swin-Unet	0.1181	0.0812	0.0527	0.0288	0.3017	0.4350	0.3118	0.2463	0.0141	0.0524	0.0073	0.0038
	Proposed SAM-B	0.6501	0.5755	0.6150	0.3830	0.6903	**0.5006**	**0.5208**	0.3941	0.5492	0.2953	0.3144	0.1564
	Proposed SAM-L	0.6503	0.5871	0.6248	0.3749	**0.7396**	0.4942	0.4698	**0.3984**	**0.6021**	**0.4452**	**0.4700**	**0.2722**
	Proposed SAM-H	**0.6518**	**0.6062**	**0.6539**	**0.4005**	0.6741	0.4660	0.4839	0.3576	0.5564	0.3965	0.4337	0.2116
Self-collected	Unet	0.5147	0.2440	0.2625	0.1155	0.552	0.2818	0.1502	0.1158	0.7400	0.6044	0.6145	0.4991
	Unet++	0.4737	0.1099	0.1151	0.0256	0.6494	0.3555	0.2806	0.2191	0.7175	0.5231	0.5306	0.4433
	Manet	0.5582	0.3265	0.3450	0.1932	**0.6954**	**0.4781**	0.2459	0.2269	0.7599	0.6399	0.6626	0.5383
	Linknet	0.5088	0.2167	0.2120	0.1111	0.5861	0.2753	0.2156	0.1485	0.7181	0.5802	0.6127	0.4739
	FPN	0.5744	0.3044	0.3136	0.2045	0.5888	0.2403	0.1751	0.1345	0.7442	0.6005	0.6262	0.5061
	PSPNet	0.5660	0.3228	0.3418	0.2052	0.5014	0.0803	0.0853	0.0528	0.7076	0.5367	0.5646	0.4440
	PAN	0.5882	0.3535	0.3659	0.2315	0.5829	0.2518	0.2142	0.1374	0.7526	0.6373	0.6682	0.5289
	DeepLabV3	0.4976	0.1363	0.1427	0.0706	0.5603	0.1833	0.1476	0.1245	0.7505	0.6179	0.6455	0.5222

(*Continued*)

Table 1. (Continued).

| Train on	Method	Test on											
		Stain				AITEX				Self-collected			
		recall	precision	F1	iou	recall	precision	F1	iou	recall	precision	F1	iou
	DeepLabV3+	0.5289	0.2833	0.3055	0.1432	0.6368	0.4050	0.1149	0.1370	0.7240	0.5667	0.5930	0.4729
	SCUnet	0.0533	0.0533	0.0952	0.0533	0.0142	0.0142	0.0267	0.0142	0.6598	0.6753	0.6416	0.5559
	Mobile–Deeplab	0.2752	0.2312	0.1856	0.1251	0.2293	0.1382	0.1199	0.0828	0.7010	0.4284	0.4881	0.4074
	Swin–Unet	0.1778	0.4129	0.1647	0.1057	0.5804	0.1164	0.1260	0.0826	0.5519	0.6629	0.5667	0.4832
	Proposed SAM-B	0.7701	0.6501	0.6485	0.5313	0.6907	0.4195	**0.3873**	0.3063	0.7738	0.6948	0.7307	0.5727
	Proposed SAM-L	**0.8462**	**0.7627**	**0.7568**	**0.6688**	0.6653	0.3776	0.3266	0.2778	**0.7855**	0.6944	0.7257	**0.5890**
	Proposed SAM-H	0.8288	0.7424	0.7447	0.6460	0.6722	0.3939	0.3790	**0.3072**	0.7791	**0.7030**	**0.7403**	0.5859
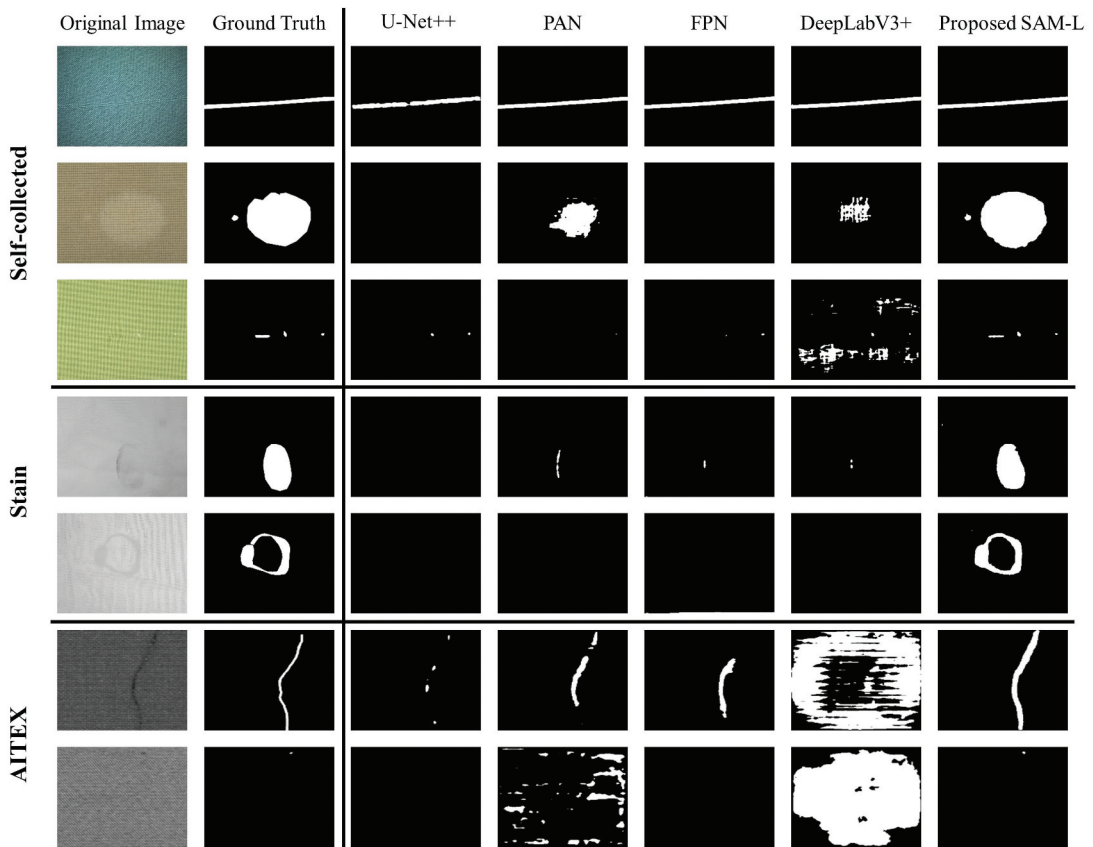
Figure 3. Comparative visualization of fabric defect segmentation results from different models. Columns from left to right respectively display: the original image, ground truth labels, predictions by U-net++, PAN, FPN, DeeplabV3+, and the proposed SAM-L model. The images showcase the superior performance of the SAM-L model in defect segmentation on three different datasets, when trained on the self-collected dataset.

Model performance of cross-dataset validation

While the diagonal entries of Table 1 illustrate the models' performance on identical training and testing datasets, the off-diagonal elements reveal the outcomes of cross-dataset validation, highlighting the models' ability to generalize when trained on one dataset and tested on another. In the cross-validation experiments, F1 score, as a comprehensive indicator of segmentation effectiveness, is the primary focus. This is because cross-dataset experiments often yield all 0 or all 1 results in predictions (i.e., no predictive capability), making some of the indicators less reliable.

When the models trained on the Stain dataset and test on the AITEX dataset, the SAM-H model achieves the best F1 score, surpassing the strongest non-SAM model, Unet++, by 7.3%. When testing on the self-collected dataset, SAM-h achieves the best F1 score, surpassing the best non-SAM model, FPN, by 4.6%. Although the SAM model still achieves the best segmentation results, the overall performance is not satisfactory. This could be due to the fact that the Stain dataset contains only the stain category of defects, leading to unsatisfactory performance on tests involving other types of defects.

When the models trained on the AITEX dataset and test on the Stain dataset, the SAM-h model achieves the best F1 score, surpassing the best non-SAM model by 43.3%. When testing on the self-collected dataset, the SAM-l model achieves the best F1 score, exceeding the best non-SAM model, DeepLabV3, by 31.4%. The results showed that the cross-database prediction of non-SAM models is unsuccessful. This is because the AITEX database has a relatively low number of training images

(approximately over 100), including only one type of defect, and has an image resolution of only 256 × 256. Such limited training data quality resulted in poor performance in cross-dataset validation for other non-SAM models, both on the stain and self-collected datasets (F1 score no more than 0.2). In contrast, the proposed SAM model demonstrates stronger segmentation performance on the stain dataset and a certain degree of segmentation capability on the self-collected dataset. These cross-validation results underscore the exceptional generalization ability of our proposed SAM model, capable of accurately segmenting noticeable defects even with limited training data quality. This also showcases the zero-shot capability of our proposed SAM model from another perspective, as the AITEX training data does not include stain-type defects.

Finally, when training on the self-collected dataset and testing on the Stain dataset, the SAM-L model achieves the best F1 score, surpassing the best non-SAM model, PAN, by 39.1%. When testing on the AITEX dataset, the SAM-b model achieves the best overall performance, exceeding the best non-SAM model, Unet++, by 10.7%. In particular, on the stain dataset, the SAM model trained on the self-collected dataset achieves a segmentation performance similar to that of non-SAM models trained on the stain dataset.

The results, as shown in Table 1, indicate that the method proposed in this study exhibits superior performance in cross-dataset validation. The visual outcomes presented in Figure 3 (rows 4 to 7) demonstrate that, when trained on the self-collected dataset, the proposed method achieves better segmentation performance on defects with slight color variations (stains in the Stain dataset) and small defects (knot in the 7th row of the AITEX dataset) compared to other models. These results highlight the robustness and precision of proposed method's segmentation capabilities when applied across diverse datasets.

Model performance on few-shot learning

Table 2 presents the performance indicators of the proposed models (SAM-B, SAM-L, SAM-H) under different few-shot learning conditions (10, 50, and 100 training samples, as well as the full training set of self-collected dataset) on the self-collected dataset. From the experiment results, it is observed that with an increase in the number of training samples from 10 to 50, the F1 scores of the SAM-B, SAM-L, and SAM-H models show significant improvement. Specifically, under the 10-shot scenario, the F1 scores are 0.4131, 0.5755, and 0.3348 for SAM-B, SAM-L, and SAM-H models respectively. However, under the 50-shot scenario, the F1 scores improve to 0.5807, 0.675, and 0.6443 for SAM-B, SAM-L, and SAM-H models respectively. Further increasing the number of training samples to 100 leads to more improvements, albeit not as significant as the previous scenario. In this 100-shot scenario, the F1 scores for SAM-B, SAM-L, and SAM-H are 0.6405, 0.6845, and 0.627 respectively. When the entire dataset (909 samples) is used for training, the performance of all models improve slightly, with F1 scores for SAM-B, SAM-L, and SAM-H being 0.7307, 0.7257, and 0.7403 respectively. Furthermore, upon examining the few-shot results, it is observed that the proposed SAM models under the 50-shot scenario are able to achieve performance levels comparable to non-SAM models trained under full-sample conditions.

Table 2. Performance metrics of the proposed SAM-B, SAM-L, and SAM-H models on the self-collected dataset under few-shot learning conditions with 10, 50, 100 training samples, and full training set scenario.

Dataset	Training Samples	Model	Accuracy	Recall	Precision	F1	Iou
Self-collected	10-shots	Proposed SAM-B	0.9852	0.6396	0.3995	0.4131	0.2876
		Proposed SAM-L	0.9838	0.6923	0.5463	0.5755	0.4109
		Proposed SAM-H	0.9343	0.6312	0.3559	0.3348	0.2618
	50-shots	Proposed SAM-B	0.9903	0.7498	0.5733	0.5807	0.4810
		Proposed SAM-L	0.9882	0.7720	0.6543	0.6750	0.5445
		Proposed SAM-H	0.9902	0.7578	0.6227	0.6443	0.5198
	100-shots	Proposed SAM-B	0.9917	0.7329	0.6079	0.6405	0.4900
		Proposed SAM-L	0.9914	0.7559	0.6507	0.6845	0.5343
		Proposed SAM-H	0.9915	0.7381	0.6006	0.6270	0.4929
	Full training	Proposed SAM-B	0.9948	0.7738	0.6948	0.7307	0.5727
		Proposed SAM-L	0.9949	0.7855	0.6944	0.7257	0.5890
		Proposed SAM-H	0.9948	0.7791	0.7030	0.7403	0.5859

Table 3. Comparison of original SAM and Proposed SAM methods across three datasets.

Method	Stain				AITEX				Self-collected			
	recall	precision	f1	iou	recall	precision	f1	iou	recall	precision	f1	iou
SAM-B (Point Prompt)	0.8355	0.4978	0.5045	0.4211	0.7318	0.1191	0.1299	0.0962	0.8641	0.2434	0.2686	0.2068
SAM-L(Point Prompt)	0.8631	0.5603	0.5999	0.5092	0.8491	0.0618	0.0779	0.0571	0.8602	0.2115	0.2503	0.1941
SAM-H(Point Prompt)	0.8767	0.6129	0.6303	0.5407	0.7863	0.1847	0.1986	0.1631	0.8499	0.2651	0.3094	0.2384
SAM-B(Box Prompt)	0.8908	0.8098	0.8279	0.7307	0.7364	0.3141	0.3151	0.2511	0.8521	0.5075	0.5755	0.4496
SAM-L(Box Prompt)	**0.9197**	0.7782	0.8178	0.7232	**0.9699**	0.2850	0.3304	0.2686	**0.8881**	0.4393	0.5255	0.4108
SAM-H(Box Prompt)	0.9176	0.7768	0.8154	0.7215	0.9604	0.2801	0.3154	0.2612	0.8836	0.4530	0.5313	0.4211
Proposed SAM-B	0.8678	0.7945	0.7978	0.7157	0.6903	**0.5006**	**0.5208**	0.3941	0.7738	0.6948	0.7307	0.5727
Proposed SAM-L	0.8911	0.8215	0.8193	0.7536	0.7396	0.4942	0.4698	**0.3984**	0.7855	0.6944	0.7257	**0.5890**
Proposed SAM-H	0.8992	**0.8356**	**0.8327**	**0.7673**	0.6741	0.4660	0.4839	0.3576	0.7791	**0.7030**	**0.7403**	0.5859

Comparison of original SAM and proposed SAM

The original Segment Anything Model (SAM) is an interactive segmentation model that requires user prompts to identify objects. Without these prompts, it cannot automatically detect defects in fabric images. To evaluate its performance in defect detection, we employed two prompting strategies: point prompts and box prompts. For point prompts, we randomly selected a point within the defect area. For box prompts, we input the coordinates of the bounding rectangle of the defect area into SAM.

Table 3 presents the performance comparison between the original SAM (with point and box prompts) and our proposed SAM across three datasets: Stain, AITEX, and Self-collected. The results demonstrate that our proposed SAM consistently outperforms the original SAM in most scenarios. On the Stain dataset, the performance of the original SAM with box prompts is comparable to our proposed SAM, with only a marginal difference of 0.48% in F1 score. However, on both the AITEX and Self-collected datasets, our proposed SAM significantly outperforms the best results of the original SAM. Specifically, on the AITEX dataset, our proposed SAM-L achieves an F1 score of 0.5208, which is 19.04% points higher than the best performance of the original SAM (box prompt). Similarly, on the Self-collected dataset, our proposed SAM-H reaches an F1 score of 0.7403, surpassing the original SAM's best performance by 16.48% points. These results demonstrate that our proposed SAM outperforms the original SAM in automated defect detection, offering improved accuracy and efficiency without the need for manual prompts.

Discussion

Segmentation performance and generalization of the proposed SAM model

By integrating the general knowledge from the original SAM with domain-specific insights into fabric defects, our proposed SAM model demonstrates exceptional performance and generalization in fabric defect segmentation. This approach proves particularly effective for identifying subtle defects and excels in cross-dataset validation scenarios.

Figure 3 presents a range of typical test results using models trained on a self-collected training set. The self-collected results shown the predictions on the test set, while the Stain and AITEX results are representative of cross-dataset validation performances.

In the examples from the self-collected dataset, three typical types of textile defects are analyzed. The first row shows a coarse latitude defect spanning the entire image, which is distinct and relatively easy to identify. For such conspicuous defects, both non-SAM models and our SAM model are capable

of effectively recognizing and accurately delineating the defect regions. However, for more challenging defects, such as those shown in the second and third rows, our SAM model outperforms the other models significantly.

The second row shows a stain that is slightly lighter in color compared to the background, where the subtle color variation poses a challenge to defect identification. The results reveal that the non-SAM models are unable to accurately recognize the complete defect area, while our SAM model can precisely detect the entire defect, even outperforming manual annotation.

The defects shown in the third row of images include coarse latitude and two small fluffs. The complexity of detecting these defects lies in their small proportion in the image and the presence of two completely different types of defects. Non-SAM models either fail to detect the defects or can only identify one of the two types (e.g., Unet++ only detects the fluffs). In contrast, the proposed SAM model can accurately identify both types of defects within a single image. Based on the observation of training and testing on the same dataset, it can be concluded that both non-SAM models and SAM models demonstrate good performance in recognizing easily-segmented defects. However, for challenging or less obvious defects, our proposed SAM model significantly outperforms non-SAM models.

Cross-dataset experiments primarily aim to test the model's generalization ability. First, two types of stains that result in surface color changes on the fabric surface were selected from the Stain dataset. The results show that non-SAM models are unable to effectively detect these defects, while our SAM model can accurately and completely identify the defect regions. Subsequently, we selected two types of defects from the AITEX dataset for detection. The first type is a prominent stain spanning the entire image, which non-SAM models can only partially detect while our SAM model can accurately and completely identify the entire defect. The second type of defect is more subtle, and all non-SAM models fail to identify this defect, leading to a high number of false detections, while the SAM model successfully identifies the defect without any false detections.

These examples underscore the superior generalization ability of our SAM model in two aspects. Firstly, the SAM model can effectively identify rare or unknown defects. For instance, in the self-collected dataset, there were only about five samples with Stain-type defects, and there were no defects similar to those in the Stain dataset, yet the SAM model could still precisely identify the defects in the Stain dataset. Secondly, the SAM model demonstrate excellent generalization capabilities to different backgrounds or imaging conditions. There are significant differences in images under different imaging conditions, which is why non-SAM models produced a high number of false detections on AITEX samples. However, the SAM model effectively resisted these influences and did not predict a high number of false detections on samples under different imaging conditions.

In summary, the importance of combining general knowledge and specialized knowledge to achieve excellent fabric defect segmentation performance is highlighted. Superior segmentation performance requires the model to understand and handle one or more specific types of defects, which necessitates a deep understanding of these defects – that's the role of specialized knowledge. However, given the diversity of types and forms of textile defects and the potential for complex situations in practical applications, relying solely on limited specialized knowledge (due to a lack of data) is insufficient. At this point, the importance of general knowledge comes into play, which includes basic skills in image processing and understanding, helping the model adapt to a changeable environment and handle a variety of defect types. Notably, in our training approach, we froze the parameters of the original SAM model, training only a small number of parameters related to fabric defects. This approach preserves the original SAM model's distribution by freezing its parameters and selectively training a limited set of parameters related to fabric defects, effectively incorporating domain-specific knowledge into the defect detection process. This strategy enriches the model with additional insights, bolstering its segmentation accuracy and generalization ability, even on small datasets. By leveraging the specialized knowledge from the fabric defect dataset alongside the broad, general knowledge embedded in the original SAM, the model adeptly handles unseen defects and adapts to variations significantly different from its training data. This balanced integration of knowledge ensures the proposed SAM model's robust performance across diverse detection scenarios.

Few-shot experiment and its practical significance

Few-shot learning experiments with the proposed SAM model reveal its impressive generalization capability from a limited dataset, underscoring its substantial practical significance, especially in specialized fields like fabric defect detection where data is often scarce.

The results from our few-shot learning experiments underscore the impressive generalization capability of our proposed SAM model. As revealed by the experimental results, with 100 training images, segmentation results that are on par with those produced by the entire 909-image training set is achieved. This not only showcases the model's effectiveness but also underscores its profound practical implications.

The process of data collection and annotation often presents challenges in terms of resource and time constraints, especially in highly specialized and intricate fields such as fabric defect detection. In such instances, models that can proficiently learn from a small number of samples become particularly indispensable. The SAM model's promising performance testifies to its edge in addressing data scarcity concerns.

The fabric industry is marked by diversity, with new types of fabrics and defects frequently making their way onto the scene, necessitating models to rapidly adapt and learn. The promising performance of the SAM model under few-shot learning conditions signifies its ability to swiftly learn and accurately identify new fabric defects within a drastically shortened training cycle, thereby amplifying its effectiveness and practicality for real-world applications.

Moreover, in scenarios related to small-scale or customized production, the ability to learn from a few samples grows even more critical. In such cases, only a very limited number of samples may be available for learning. The SAM model's capacity to rapidly learn and deliver exceptional segmentation performance in such environments allows it to better cater to these real-world application scenarios.

Performance comparison between original SAM and proposed SAM

This study compared the performance of the original SAM and the proposed SAM across three different datasets (Stain, AITEX, and Self-collected). The results reveal significant differences in the performance of the two methods across these datasets.

On the Stain dataset, the original SAM (especially when using box prompts) achieved segmentation performance similar to the proposed SAM. This is primarily because the defects in the Stain dataset are mostly represented as black stains on white fabric, a relatively simple presentation that shares similarities with many elements in natural images. Consequently, when provided with explicit box prompts, the original SAM can effectively identify and segment these types of defects. However, on the AITEX and Self-collected datasets, even with strong prompts (box), the original SAM failed to achieve segmentation performance comparable to the proposed SAM. These two datasets contain more fabric-specific defect types, such as thick yarn, broken yarn, knots, and weft density abnormalities. These specialized fabric defects differ significantly in form and characteristics from common objects in natural images. The original SAM model, lacking specialized knowledge of fabric defects, struggles to accurately identify and segment these particular types of flaws.

The advantage of the proposed SAM lies in its ability to automatically detect various types of fabric defects without manual prompts. This capability not only improves detection accuracy but also greatly increases the efficiency of the detection process. More importantly, the proposed SAM demonstrates good adaptability to different types of datasets, exhibiting higher stability when dealing with complex and diverse fabric defects. These findings highlight the importance and innovation of this research. By combining the SAM model with specialized knowledge of fabric defects, our method can better address the fabric quality control needs in real industrial environments, providing a more reliable and efficient solution for automated fabric inspection.

Partial parameter training: efficiency and cost-effectiveness

Partial parameter training in the proposed SAM model emerges as a highly efficient and cost-effective strategy, optimizing computational resource use while maintaining robust model performance in fabric defect detection.

As illustrated in the Table 4, only a minor proportion of the total parameters are trainable, indicating the efficient utilization of computational resources.

For the original SAM, training all parameters demand extensive computational resources and time. However, by focusing on just a subset of the model's parameters, the computational resources needed can be significantly reduced, thereby significantly boosting the efficiency of model training. This approach is exceptionally beneficial in research or production environments where resources are constrained.

Training only a selected subset of parameters within the fabric defect detection model yields commendable results. This strategy effectively leverages the preexisting general knowledge embedded within the majority of pre-trained parameters. The specific adaptation to the task is handled by targeted training of a minimal subset of parameters. This methodology not only enhances training efficiency but also preserves the model's generalization ability. Consequently, the model is equipped to maintain superior performance, even when encountering previously unseen types of fabric defects.

Moreover, the tactic of training just part of the model's parameters can serve as an effective guard against overfitting. In our experiments, given the relatively limited number of training samples, overfitting is a critical concern. By limiting the number of parameters to be trained, model performance can be preserved while circumventing the loss of predictive ability for new samples due to overfitting the training data.

Annotation correction and assistance

Due to the robust generalization ability of the SAM model, it has a certain tolerance for errors in data annotation and will not overfit erroneous labels. As shown in Figure 4, the segmentation results of SAM are superior to manual annotations, with smoother edges.

This characteristic bears significant value for practical applications. Firstly, manual annotation processes are inevitably prone to mistakes, and the error-tolerance of the SAM effectively alleviates the impact of such mislabels on model learning. Furthermore, the SAM model can be employed to assist in manual annotation, automatically conducting preliminary defect segmentation and generating annotation suggestions, which significantly saves time and effort in manual annotation. At the same time, given that the segmentation results of the SAM model have smooth edges, for edge cases difficult to discern by the human eye, the suggestions from the SAM model can provide valuable references. Consequently, the SAM model serves not only as a tool for defect detection but also as a potent annotation tool, improving annotation efficiency, reducing annotation errors, and thereby enhancing the overall efficiency and accuracy of fabric defect detection.

Limitations and future works

This study presents an effective method for utilizing Large Visual Models (LVMs) in fabric defect segmentation tasks. Although our approach has achieved significant results, there are still aspects that can be further optimized. Firstly, our method is essentially an efficient way of

Table 4. Total and trainable parameters of the proposed SAM models.

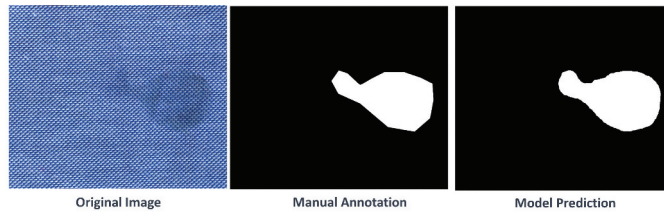| Model	Total Parameters	Trainable Parameters
Proposed SAM-B	91233774	657442
Proposed SAM-L	309300462	903202
Proposed SAM-H	637515758	1165346

Figure 4. Comparison of manual annotation and model prediction.

transferring a natural image foundation model to the task of fabric defect detection. While this approach can achieve better segmentation performance and generalization ability with limited training data, SAM itself is not a foundation model specifically designed to address several core issues in fabric defect detection. For instance, in dealing with extremely small defects, highly complex textured backgrounds, extremely rare defect types, and performance under different backgrounds and imaging conditions, the performance of a foundation model trained on natural images may still have room for improvement. However, it is worth noting that our experimental results demonstrate that this method performs excellently in most common fabric defect detection tasks, laying a solid foundation for addressing these more challenging issues. Secondly, although our method only requires training a small portion of the model's parameters, the proposed SAM still retains most of the original model structure, including versions with 90 million, 300 million, and 600 million parameters. This means that the model may occupy substantial GPU memory during operation, and inference time may be longer compared to some lightweight models. In some industrial environments with extremely high real-time requirements, this could pose a challenge.

For future work, we plan to establish a fabric foundation model by injecting more general knowledge of the fabric domain into the model. This may necessitate pre-training on a large-scale fabric image dataset to capture the generic knowledge of fabric textures, colors, lighting, etc. Moreover, we may design a fabric-specific interactive segmentation, allowing users to perform fabric defect detection and annotation more flexibly. To address the computational demands and inference time challenges, we will explore model compression and acceleration techniques, such as knowledge distillation, pruning, or quantization, to reduce model size and inference time while maintaining high accuracy. This optimization will aim to make our approach more suitable for real-time industrial applications with high-speed requirements.

Conclusion

This study focuses on improving the fabric defect segmentation capacity of the pre-trained SAM model. This is achieved through the integration of defect-specific knowledge via a tailored set of trainable parameters. The proposed SAM model demonstrates remarkable resilience, streamlined training, and superior performance across a range of scenarios, such as few-shot learning and cross-dataset validation. Moreover, it presents a high tolerance for annotation errors and offers assistance in annotation correction. These advancements elevate the accuracy and efficiency of fabric defect detection. This pioneering approach paves the way for considerable advancements in textile quality control, forecasting substantial resource optimization and reduction in required efforts.

Disclosure statement

References

Arora, P., and M. Hanmandlu. 2022. "Detection of Defects in Fabrics Using Information Set Features in Comparison with Deep Learning Approaches." *The Journal of the Textile Institute* 113 (2): 266–272. https://doi.org/10.1080/00405000.2020.1870326.

Bai, Z., and J. Jing. 2023. "Mobile-Deeplab: A Lightweight Pixel Segmentation-Based Method for Fabric Defect Detection." *Journal of Intelligent Manufacturing* 35 (7): 3315–3330. https://doi.org/10.1007/s10845-023-02205-1.

Boluki, M., and F. Mohanna. 2021. "Inspection of Textile Fabrics Based on the Optimal Gabor Filter." *Signal, Image and Video Processing* 15 (7): 1617–1625. https://doi.org/10.1007/s11760-021-01897-3.

Canatar, A., B. Bordelon, and C. Pehlevan. 2021. "Spectral Bias and Task-Model Alignment Explain Generalization in Kernel Regression and Infinitely Wide Neural Networks." *Nature Communications* 12 (1): 2914. https://doi.org/10.1038/s41467-021-23103-1.

Cen, J., Y. Wu, K. Wang, X. Li, J. Yang, Y. Pei, L. Kong, Z. Liu, and Q. Chen. 2023. "Sad: Segment Any Rgbd." *arXiv preprint arXiv:2305.14207*.

Chaurasia, A., and E. Culurciello. 2017. "Linknet: Exploiting Encoder Representations for Efficient Semantic Segmentation." In *2017 IEEE Visual Communications and Image Processing (VCIP)*, St. Petersburg, FL, USA, 1–4. IEEE.

Chen, L.-C., G. Papandreou, F. Schroff, and H. Adam. 2017. "Rethinking Atrous Convolution for Semantic Image Segmentation." *arXiv preprint arXiv:1706.05587*.

Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. 2018. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation." In *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, edited by V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Vol. 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_49.

Chen, T., M. Li, M. Li, Y. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang. 2015. "Mxnet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems." *arXiv preprint arXiv:1512.01274*.

Cheng, L., J. Yi, A. Chen, and Y. Zhang. 2023. "Fabric Defect Detection Based on Separate Convolutional Unet." *Multimedia Tools & Applications* 82 (2): 3101–3122. https://doi.org/10.1007/s11042-022-13568-7.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 248–255. Ieee.

Deng, R., C. Cui, Q. Liu, T. Yao, L. W. Remedios, S. Bao, B. A. Landman, et al. 2023. "Segment Anything Model (Sam) for Digital Pathology: Assess Zero-Shot Segmentation on Whole Slide Imaging." *arXiv preprint arXiv:2304.04155*.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, et al. 2020. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *arXiv preprint arXiv:2010.11929*.

Fan, T., G. Wang, Y. Li, and H. Wang. 2020. "Ma-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation." *Institute of Electrical and Electronics Engineers Access* 8:179 656–179 665. https://doi.org/10.1109/ACCESS.2020.3025372.

Han, X., Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, et al. 2021. "Pre-Trained Models: Past, Present and Future." *AI Open* 2:225–250. https://doi.org/10.1016/j.aiopen.2021.08.002.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 770–778.

Hu, E. J., Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. 2021. "Lora: Low-Rank Adaptation of Large Language Models." *arXiv preprint arXiv:2106.09685*.

Huang, Y., J. Jing, and Z. Wang. 2021. "Fabric Defect Segmentation Method Based on Deep Learning." *IEEE Transactions on Instrumentation and Measurement* 70:1–15. https://doi.org/10.1109/TIM.2020.3047190.

Ji, W., J. Li, Q. Bi, T. Li, W. Li, and L. Cheng. 2023. "Correction To: Segment Anything is Not Always Perfect: An Investigation of SAM on Different Real-World Applications." *Machine Intelligence Research. arXiv preprint arXiv:2304.05750.* https://doi.org/10.1007/s11633-024-1526-0.

Jing, J., Z. Wang, M. Rätsch, and H. Zhang. 2022. "Mobile-Unet: An Efficient Convolutional Neural Network for Fabric Defect Detection." *Textile Research Journal* 92 (1–2): 30–42. https://doi.org/10.1177/0040517520928604.

Jing, Y., X. Wang, and D. Tao. 2023. "Segment Anything in Non-Euclidean Domains: Challenges and Opportunities." *arXiv preprint arXiv:2304.11595*.

Kahraman, Y., and A. Durmuşoğlu. 2023. "Deep Learning-Based Fabric Defect Detection: A Review." *Textile Research Journal* 93 (5–6): 1485–1503. https://doi.org/10.1177/00405175221130773.

Kirillov, A., K. He, R. Girshick, and P. Dollár. 2017. "A Unified Architecture for Instance and Semantic Segmentation."

Kirillov, A., E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, et al. 2023. "Segment Anything." In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/ICCV51070.2023.00371.

Kopaczka, M., M. Saggiomo, M. Güttler, K. Kielholz, and D. Merhof. 2019. "Detection and Classification of Faulty Weft Threads Using Both Feature-Based and Deep Convolutional Machine Learning Methods." In *Pattern Recognition Applications and Methods: 7th International Conference, ICPRAM 2018, Funchal, Madeira, Portugal, January 16-18, 2018, Revised Selected Papers 7*, Funchal, Madeira, Portugal, 141–163. Springer.

Koulali, I., and M. T. Eskil. 2021. "Unsupervised Textile Defect Detection Using Convolutional Neural Networks." *Applied Soft Computing* 113:107913. https://doi.org/10.1016/j.asoc.2021.107913.

Larsen, R., T. L. Villadsen, J. K. Mathiesen, K. M. Ø. Jensen, and E. D. Boejesen. 2023. "Np-Sam: Implementing the Segment Anything Model for Easy Nanoparticle Segmentation in Electron Microscopy Images." *ChemRxiv*. https://doi.org/10.26434/chemrxiv-2023-k73qz-v2.

Li, B., Y. Zou, R. Zhu, W. Yao, J. Wang, S. Wan, and V. Menon. 2022. "Fabric Defect Segmentation System Based on a Lightweight Gan for Industrial Internet of Things." *Wireless Communications and Mobile Computing* 2022:1–17. https://doi.org/10.1155/2022/9680519.

Li, C., J. Li, Y. Li, L. He, X. Fu, J. Chen, and X. Zhou. 2021. "Fabric Defect Detection in Textile Manufacturing: A Survey of the State of the Art." *Security and Communication Networks* 2021:1–13. https://doi.org/10.1155/2021/9948808.

Li, H., P. Xiong, J. An, and L. Wang. 2018. "Pyramid Attention Network for Semantic Segmentation." *arXiv preprint arXiv:1805.10180*.

Li, Y., W. Zhao, and J. Pan. 2016. "Deformable Patterned Fabric Defect Detection with Fisher Criterion-Based Deep Learning." *IEEE Transactions on Automation Science and Engineering* 14 (2): 1256–1264. https://doi.org/10.1109/TASE.2016.2520955.

Liu, J., C. Wang, H. Su, B. Du, and D. Tao. 2019. "Multistage Gan for Fabric Defect Detection." *IEEE Transactions on Image Processing* 29:3388–3400. https://doi.org/10.1109/TIP.2019.2959741.

Liu, Z., J. Wang, C. Li, B. Li, and R. Yang. 2019. "Fabric Defect Detection Using Fully Convolutional Network with Attention Mechanism." In *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*, Beijing China, 134–140.

Loshchilov, I., and F. Hutter. 2017. "Decoupled Weight Decay Regularization." *arXiv preprint arXiv:1711.05101*.

Ma, J., Y. He, F. Li, L. Han, C. You, and B. Wang. 2023. "Segment Anything in Medical Images." *Nature Communications* 15:654. https://doi.org/10.1038/s41467-024-44824-z.

Mo, S., and Y. Tian. 2023. "Av-Sam: Segment Anything Model Meets Audio-Visual Localization and Segmentation." *arXiv preprint arXiv:2305.01836*.

Ngan, H. Y., G. K. Pang, and N. H. Yung. 2011. "Automated Fabric Defect Detection—A Review." *Image and Vision Computing* 29 (7): 442–458. https://doi.org/10.1016/j.imavis.2011.02.002.

Niu, S., B. Li, X. Wang, S. He, and Y. Peng. 2022. "Defect Attention Template Generation Cyclegan for Weakly Supervised Surface Defect Segmentation." *Pattern Recognition* 123:108396. https://doi.org/10.1016/j.patcog.2021.108396.

Ouyang, W., B. Xu, J. Hou, and X. Yuan. 2019. "Fabric Defect Detection Using Activation Layer Embedded Convolutional Neural Network." *Institute of Electrical and Electronics Engineers Access* 7:70130–70140. https://doi.org/10.1109/ACCESS.2019.2913620.

Pathirana, P. 2020. "Fabric Stain Dataset." Accessed December 10, 2021. https://www.kaggle.com/priemshpathirana/fabric-stain-dataset.

Peng, P., Y. Wang, C. Hao, Z. Zhu, T. Liu, and W. Zhou. 2020. "Automatic fabric defect detection method using pran-net." *Applied Sciences* 10 (23): 8434. https://doi.org/10.3390/app10238434.

Rong-Qiang, L., L. Ming-Hui, S. Jia-Chen, and L. Yi-Bin. 2021. "Fabric Defect Detection Method Based on Improved U-Net." *Journal of Physics: Conference Series* 1948 (1). 012160. IOP Publishing. https://doi.org/10.1088/1742-6596/1948/1/012160.

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Munich, Germany, 234–241. Springer.

Shao, L., E. Zhang, Q. Ma, and M. Li. 2022. "Pixel-Wise Semisupervised Fabric Defect Detection Method Combined with Multitask Mean Teacher." *IEEE Transactions on Instrumentation and Measurement* 71:1–11. https://doi.org/10.1109/TIM.2022.3162286.

Silvestre-Blanes, J., T. Albero-Albero, I. Miralles, R. Pérez-Llorens, and J. Moreno. 2019. "A Public Fabric Database for Defect Detection Methods and Results." *AUTEX Research Journal* 19 (4): 363–374. https://doi.org/10.2478/aut-2019-0035.

Sun, G., Z. Zhou, Y. Gao, Y. Xu, L. Xu, and S. Lin. 2019. "A Fast Fabric Defect Detection Framework for Multi-Layer Convolutional Neural Network Based on Histogram Back-Projection." *IEICE Transactions on Information and Systems* E102.D (12): 2504–2514. https://doi.org/10.1587/transinf.2019EDP7092.

Tang, L., H. Xiao, and B. Li. 2023. "Can Sam Segment Anything? When Sam Meets Camouflaged Object Detection." *arXiv preprint arXiv:2304.04709*.

Tian, H., and F. Li. 2019. "Autoencoder-Based Fabric Defect Detection with Cross-Patch Similarity." In *2019 16th International Conference on Machine Vision Applications (MVA)*, Tokyo, Japan, 1–6. IEEE.

Wu, J., R. Fu, H. Fang, Y. Liu, Z. Wang, Y. Xu, Y. Jin, and T. Arbel. 2023. "Medical Sam Adapter: Adapting Segment Anything Model for Medical Image Segmentation." *arXiv preprint arXiv:2304.12620*.

Xu, H., C. Liu, S. Duan, L. Ren, G. Cheng, and B. Hao. 2023. "A Fabric Defect Segmentation Model Based on Improved Swin-Unet with Gabor Filter." *Applied Sciences* 13 (20): 11386. https://doi.org/10.3390/app132011386.

Ying, X. 2019. "An Overview of Overfitting and Its Solutions." *Journal of Physics: Conference Series* 1168. 022022. IOP Publishing. https://doi.org/10.1088/1742-6596/1168/2/022022.

Yu, T., R. Feng, R. Feng, J. Liu, X. Jin, W. Zeng, and Z. Chen. 2023. "Inpaint Anything: Segment Anything Meets Image Inpainting." *arXiv preprint arXiv:2304.06790.*

Zhang, R., Z. Jiang, Z. Guo, S. Yan, J. Pan, H. Dong, P. Gao, and H. Li. 2023. "Personalize Segment Anything Model with One Shot." *arXiv preprint arXiv:2305.03048.*

Zhao, H., J. Shi, X. Qi, X. Wang, and J. Jia. 2017. "Pyramid Scene Parsing Network." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2881–2890.

Zhao, S., L. Yin, J. Zhang, J. Wang, and R. Zhong. 2020. "Real-Time Fabric Defect Detection Based on Multi-Scale Convolutional Neural Network." *IET Collaborative Intelligent Manufacturing* 2 (4): 189–196. https://doi.org/10.1049/iet-cim.2020.0062.

Zhou, Z., M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang. 2018. "Unet++: A Nested U-Net Architecture for Medical Image Segmentation." In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings*, Granada, Spain, 3–11. Springer.

Zhou, Z., Z. Wu, R. Boutteau, F. Yang, and D. Ginhac. 2023. "Dsec-Mos: Segment Any Moving Object with Moving Ego Vehicle." *arXiv preprint arXiv:2305.00126.*

Zhu, Z., G. Han, G. Jia, and L. Shu. 2020. "Modified Densenet for Automatic Fabric Defect Detection with Edge Computing for Minimizing Latency." *IEEE Internet of Things Journal* 7 (10): 9623–9636. https://doi.org/10.1109/JIOT.2020.2983050.