Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Economics and Finance 00 (2012) 000–000

Procedia
Economics and Finance

www.elsevier.com/locate/procedia

Memory Linked Knowledge Domain Transfer Few-shot Learning for Thermography Nondestructive Evaluation System

Jianguo Xue[a], Bin Gao[a]*, Guohao Liu[a], Yuming Zhang[a], Wai Lok Woo[b], Yang Yang[c], Yongjie Yu[c],

[a] *the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[b]*Department of Computer and Information Sciences, NE1 7RU, Northumbria University, England, UK*
.[c]*Chengdu aircraft industry Co., Ltd. Chengdu 611731, China*

Abstract

Infrared NDT techniques are suitable for detecting near-surface and sub-surface defects in composite materials. However, significant detection challenges remain while data volume is quantitatively limited for training weak contrast between defective and non-defective regions. In this paper, we propose an end-to-end memory linked knowledge domain with transfer few-shot learning segmentation network with a wider general model that solves the problems of insufficient data support. The proposed method is shown to attain high detection performance for multiple types of specimens by using cross domain adaption. By improving the support branch, a foreground-background joint guidance linked with memory bank is proposed. This provides better guidance information for the query branch. In addition, the intra-class and inter-class loss metric distance between defective and non-defective features is increased for better segmentation performance. In order to enable interpretability of the model, robust experiments have been conducted to detect the inner debond on multiple carbon fiber reinforced polymer (CFRP) composites. A comparative analysis is presented and compared with the state-of-the-art machine learning algorithms.

"Keywords: Optically Pulsed Infrared Nondestructive Testing, Deep Learning, Few-shot Learning, Image Semantic Segmentation;"

1. INTRODUCTION

Deep convolutional neural networks have been widely applied to visual understanding applications, e.g., image classification, object detection and semantic segmentation. Supervised segmentation is the task of classifying each pixel that is used an image into different areas. Most existing semantic segmentation methods are based on fully convolutional networks (FCNs) and its variants since transposed convolution is used to complete dense predictions tasks instead of fully connected layers. Chen et al. proposed DeepLab[1] based on FCN[2][3]. It employs conditional random fields (CRF) to find relationships between close pixels. Similar

* Corresponding author: Bin Gao.
E-mail address: bin_gao@uestc.edu.cn.

approaches are including SegNet[4], RefineNet[5], PSPNet[6]. Ronneberger et al. proposed UNet[7] based on the encoder-decoder architecture and skip connections which has excellent segmentation performance in the field of medicine and defect detection. Overall, these studies contribute diverse methodologies, such as the pyramid pooling module[8], dilated convolution [9], deformable convolution[10], multi-scale feature aggregation[11], and attention mechanisms[12]. Although these methods have achieved good results, they require huge training resources and strong pixel-level marking of ground truths. In addition, the trained model cannot recognize new classes that do not exist in the training set in evaluation. These problems occur not only in semantic segmentation, but also in the entire field of visual comprehension.

To resolve the generalization problem, few-shot learning[14] methods have been proposed to treat the problem of image classification. At present, there are mainly two categories: metric-learning and meta-learning. Metric-learning[15] is also called similarity learning, that is, learning an embedding space where all data is converted into a feature vector to achieve the goal of maximizing the inter-class variations and minimizing the intra-class variations. Meta learning[16] increases the ability of the learner to normalize multiple tasks by defining specific optimization, loss functions and other methods. In summary, few-shot learning improves the generalization ability of the model. After training, only one or few annotated examples are required to learn to predict new classes. With the deepening of research, scholars have expanded the few-shot learning methods to the field of semantic segmentation. This is a more meaningful and challenging task. On one hand, semantic segmentation requires predicting labels for each pixel rather than individual labels for the entire image. On other hand, few-shot learning methods solves the problem that semantic segmentation labels need a lot of manual time marking. Shaban et al. [16] first proposed a dual-branch network for few-shot segmentation. The main strength of co-FCN[17] is that only few sparse labelling of the original image can achieve pixel-level segmentation of the target. SG-One[18] proposed masked average pooling to extract the feature vector of the target in support set, and cosine similarity is used to measure the distance between the representation vector of query set and the representation vector of support set to guide the segmentation of query set. PANet [19] introduced prototypes alignment regularization, making full use of support knowledge. CANet[20] expanded attention mechanism effectively fuses information from multiple support examples in k-shot settings, which is performed better than the unlearnable fusion method for single result. PGNet[21] used the graph attention mechanism and a pyramid-like structure on this basis, which model image regions of different sizes as graph nodes to obtain correspondences at different semantic levels. PMMs[22] used EM algorithm to estimate multiple prototypes which improved the semantic representation ability. Yang et al. proposed two modules: superpixel guided clustering (SGC) and guided prototype assignment (GPA). In particular, the generated prototypes become content adaptive and spatial aware to adapt to changes in the proportion as well as shape of objects. Current few-shot segmentation networks usually extract features from query images and support images for feature matching. Sun et al.[23] proposed a novel few-shot learning method with meta-transfer learning (MTL) that aims to adapt a deep neural network for few shot learning tasks. The MTL contributes to fast convergence and high accuracy in these tasks. Wu et al. [24] proposed the construction of a few-shot transfer learning method for diagnosing samples in variable conditions, using meta-learning. The paper presents seven different few-shot transfer learning methods, all based on a unified 1D convolution network, for diagnosing three different datasets.

Non-Destructive Testing (NDT) is an important application field of computer vision for defects detection. NDT and health tracing of materials are essential for the safety of equipment and personnel. Optical pulsed thermography (OPT)[25] system is one of NDT and evaluation methods. It has the advantages of non-contact, large detection area and high speed. The OPT system can detect near-surface and subsurface defects in many materials, such as glass fiber, coated materials, carbon fiber reinforced polymer (CFRP)[26] materials, etc. Many methods are proposed to process infrared thermal image tensor[27] data generated by OPT systems. Typical unsupervised algorithms are used to improve signal-to-noise ratio and reduce data dimensions to make it easier to detect defects[28], such as PCA[29][30], PPT[31], TSR[32], and matrix decomposition algorithms[33][34]. In addition, deep learning algorithm[35] have been employed to detect and calculate defect area. Most of these algorithms are improved by mainstream object detection and semantic segmentation

methods. These algorithms can only be used for a certain material, or even a certain type of defect. The shape of the material and the size as well as depth of the defect will affect the performance of the algorithm due to the thermal conductivity, and the reflection degree on the light source excitation. In addition, the infrared thermal image tensor data have problems of low resolution, high noise and uneven heating. In practical applications, the amount of data available for each material is much smaller than the field of semantic segmentation due to the limitation of the number of specimen and natural defects. Guaragnella et al. [40] proposed a method for detecting defects in composite materials using a deep learning algorithm and lockin thermography. The algorithm is trained on a high-variance dataset obtained using five different heat source configurations. Experimental results show that this methodology is effective in automatically detecting defects in samples with known defects. Liu et al. [41] proposed a Mask Region-based Convolution NN (Mask-RCNN) for detecting and segmenting defects using a single thermal frame. The results had shown that this method outperformed several state-of-the-art post-processing methods in IRT, particularly in identifying detailed defect corners and edges.

To address the above issues, this paper proposes an end-to-end few-shot learning segmentation network with a model that addresses the lack of data volume and generalization performance of multiple types of specimens. By using the foreground-background joint guidance, the network can provide informed guidance information for the query branch (in the few shot learning) such that it increases the metric distance between defective and non-defective features for better segmentation of easy-to-miss defects. To summarize, the contributions of the paper are as follows:

(i) Proposing a domain-adaptive few-shot learning defect segmentation algorithm to reduce the class differences between support set and query sets, thus reduces the difficulty of extracting common information between few-shot network learning support set and query sets. This subsequently enhances the convergence speed of few-shot learning network and improves the ability of mining difficult samples. Domain-aligned support images can provide better guidance information.

(ii) Proposing a query branch structure with joint foreground and background guidance whose purpose is to provide better informed guidance information. To this effect, a memory guidance module is introduced to record multiple memory prototypes and to provide more detailed information for the query branch. Through the update of the memory module, the provided background guidance information will not be limited to the current support image since the memory module can memorize multiple types of backgrounds of historical support images. The joint foreground-background guidance increases the metric distance between foreground and background, and reduces missed and false detections.

(iii) Proposing intra-class and inter-class losses for the memory module so that the limited memory module can store more relevant information and to ensure the diversity and accuracy of memory prototypes. The goal of small intra-class distance and large inter-class distance of memory prototypes in the memory module is achieved by optimizing the loss function.

The effectiveness and robustness of the proposed algorithm are analyzed and verified from four aspects: visualization of defect detection results, objective evaluation of Intersection over Union (IOU) and F-score evaluation indexes, ablation experiments, and comparison with similar state-of-the-art methods.

2. METHODOLOGY

2.1. Data Preprocessing

As the model runs on a portable device, data pre-processing is required to reduce the data dimensionality for the infrared thermal imaging video. For the thermographic sequences $\mathbf{A} \in \mathrm{R}^{M \times N \times F}$ where ($M \times N$) is the size of the frame and F represents the number of frames of the sequences. The infrared thermal imaging video reduces the dimension with a sequence PCA algorithm or other tensor decomposition algorithm to obtain K infrared thermal images of high signal-to-noise ratio.

2.2. Guidance Network Construction

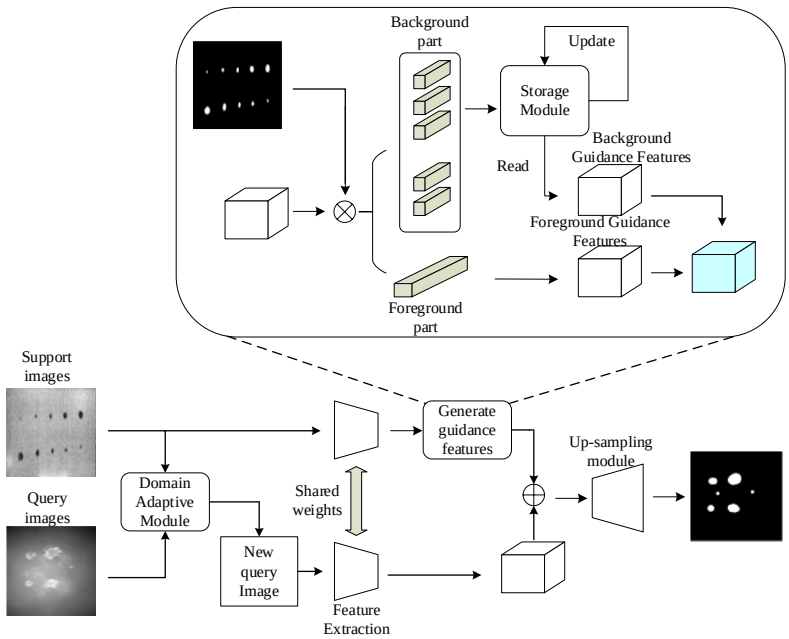


Fig. 1 Overall block diagram of the few-shot semantic segmentation algorithm

Existing few-shot learning segmentation approaches mostly ignore the background and past support image information. Thus, this section is dedicated to developing and implementing the storage update of background information through the memory module, and implements local feature storage of historical background information. Global foreground features and multiple background features integrate the contextual information, background information, current information and historical information. On the feature extraction of the support image and query image, this paper uses the ResNet-50 as a backbone network of the feature extractor, the backbone network parameter is the pre-trained parameters on the ImageNet, which remains unchanged during the training process. In the past CNN feature visualization, it has shown that low-level features contain typically low semantic information such as edge and color. High-level features are global information, with larger receptive fields, including information such as object categories. In few-shot learning, the model needs to adapt to the new unlearned categories. Therefore, the requirement of a high-level feature corresponding to an unknown new category is that it should not be obtained by learning during the training. Hence, the network in the proposed model is more concerned with the intermediate layer characteristics of unknown new categories. The proposed model selects the semantic information in the middle layer to compare the differences between the support feature map and the query feature map.

Most of the current research uses only the foreground feature to generate guidance information, that is, only the defective area information guides the query branch. The network can only learn the common part of defect information. The global foreground feature will lose the details of the defects while some defects with low contrast than the background are more difficult to identify, and these defects are easy to cause leak detection rates due to infrared noise. Of the unknown type of defect image, the network cannot utilize the difference information of the background and defective part due to the lack of background information guidance. Therefore, a new memory module is proposed where the information of the background is stored in the memory module while the query branch is guided by the joint instructions of the foreground and historical background.

In previous studies, some networks attempted to store long-term dependencies in sequence data. For example, LSTM proposes an internal memory unit but for few-shot learning tasks, LSTM has limited memory performance. In order to overcome this limit, recent researchers put forward memory networks.

For the defective area information extracted by the support image, the proposed method uses the masked average pooling to compress the feature into a feature vector, with a feature vector corresponding to the target category to represent the defective information of the support image.

We use the memory module to characterize historical background information and combine it with the global foreground feature[36]. The idea is that the key value is used to constitute a memory module, using different loss functions to update memory. This records the various modes of the background data into a single vector of the memory module and treats each memory vector as a prototype feature.

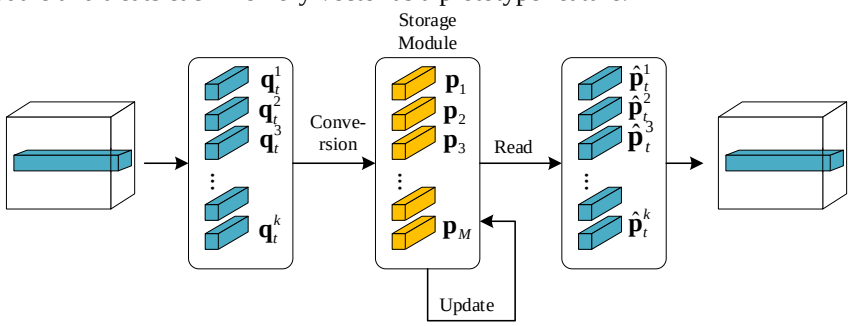


Fig. 2 Storage module data change process

Fig 2 shows the storage module data variation process. The blue portion in the figure is a unit of the support image feature map, the yellow portion is a memory prototype, and each memory prototype is a vector.

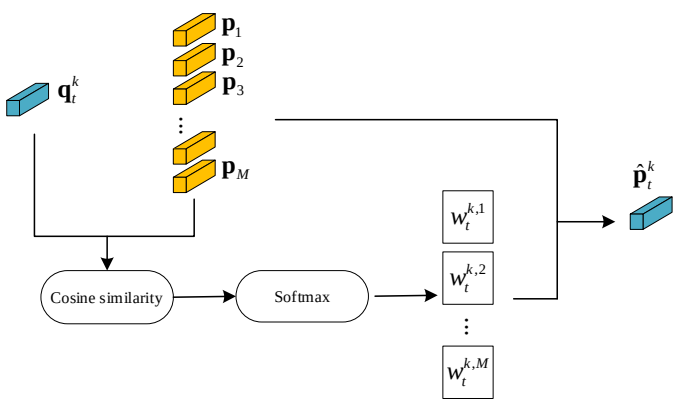


Fig.3 Memory module reading process

As shown in Fig. 3, a support feature map is generated by feature extraction, and $\mathbf{q}_t^k$ is a background feature in the feature map Fs. $\mathbf{p}_m$ is the M few-shot prototype of the memory module (m =1,2,3…, M), each $\mathbf{p}_m$ calculation of cosine similarity between $\mathbf{q}_t^k$. This requires matching the support image background features with different categories of prototypes, i.e., the similarities between each of the support image background features and different memory prototypes are calculated. Finally, we apply the softmax function to get the matching probability value $w_t^{k,m}$.

$$w_t^{k,m} = \frac{\exp\left(\left(\mathbf{p}_m\right)^T \mathbf{q}_t^k\right)}{\sum_{m'=1}^{M} \exp\left(\left(\mathbf{p}_{m'}\right)^T \mathbf{q}_t^k\right)} \quad (1)$$

where m is the number of memory prototypes, with a total of M memory prototypes; $\mathbf{q}_t^k$ represents each background characteristic of the support feature map.

Since the data structure of the memory prototype is a key-value pair, it is not possible to directly use the subsequent up-sampling operation. Thus, it is necessary to convert the memory prototype into a feature map by reading operations. By using $\mathbf{p}_m$, we determine the probability-weighted average to calculate new feature $\hat{\mathbf{p}}_t^k$:

$$\hat{\mathbf{p}}_t^k = \sum_{m'=1}^{M} w_t^{k,m'} \mathbf{p}_{m'} \quad (2)$$

The proposed method uses all memory items to reconstruct new features $\hat{\mathbf{p}}_t^k$, the purpose is to enable the model to learn the background mode of different types of test pieces. By concatenating the $\hat{\mathbf{p}}_t^k$ and the foreground feature map generated by intensive comparison module along the channel dimension, we can form the combined foreground-background guidance input into the decoder.

The way for update is critical to memory modules since it memorizes multi-class background features. For each memory prototype, like a flow of reading operations, it computes the cosine similarity between each Q and P. In addition, this is matching the support image background features to different categories of memory prototypes, i.e., the similarities between each item and different prototypes are calculated. Thus, the matching probability value $w_t^{k,m}$ is obtained by the softmax function. U_t^m shows the collection of $w_t^{k,m}$ corresponding to the m memory item $\mathbf{p}_m$. As shown in Fig 4, U_t^m represents the index collection of the $w_t^{k,m}$, $w_t^{'k,m}$ corresponding to the yellow memory item $\mathbf{p}_m$.

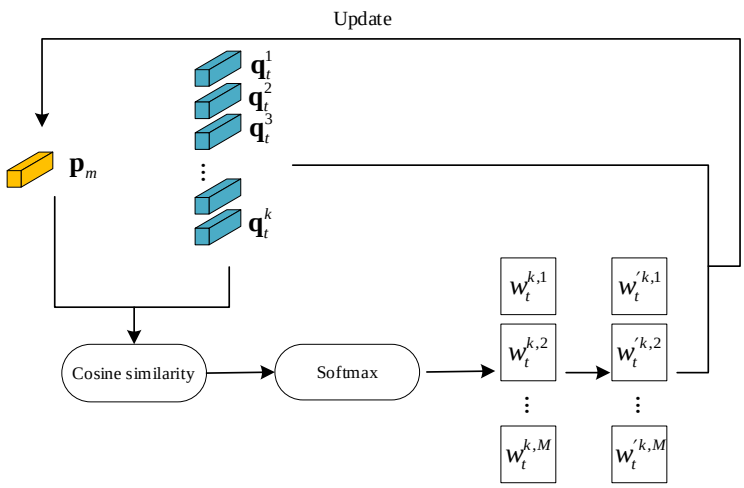


Fig.4 Memory module update process

The term $w_t^{'k,m}$ is a normalized matching probability as determined in the following Eq. 3:

$$w_t^{'k,m} = \frac{w_t^{k,m}}{\max\limits_{k' \in U_t^m} w_t^{k',m}} \quad (3)$$

$$\mathbf{p}_m \leftarrow f\left(\mathbf{p}_m + \sum_{k \in U_t^m} w_t^{'k,m} \mathbf{q}_t^k\right) \tag{4}$$

The update of the mth memory item $\mathbf{p}_m$ is given in Eq. 4 with $f(\bullet)$ representing the L_2 regularization. The purpose is to update $\mathbf{q}_t^k$ information similar to the $\mathbf{p}_m$ to the memory item. Unlike the reading flow of the memory module, the update of the memory module is only processed in training.

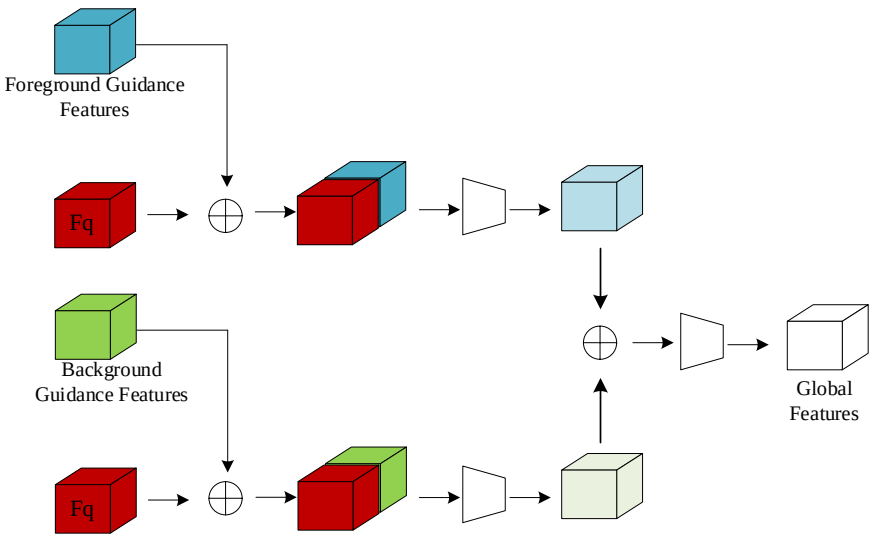


Fig. 5 Guidance feature calculation global feature process

From the above, both foreground and background have obtained foreground guidance feature and background guidance features, respectively. The foreground portion is connected to all spatial positions after obtaining a global feature vector from the support set. All spatial locations in the query branch are compared to the global feature vector.

The background prototype is updated in the memory module, and the background guidance feature of the same space size is obtained by the reading operation. Connect with the spatial dimension of F, as shown in Fig. 5.

The convolution operation in Fig. 5 is of size 3×3. Once global feature is obtained, the predicted results can be obtained by the upsampling decoder module. The decoder uses the Atrous Spatial Pyramid Pool (ASPP) to capture multi-scale information.

2.3. Proposed Loss function

In order to train the few-shot learning model of the memory module, the loss function of the memory module needs to be designed. To achieve a small distance of the same category, the distance between different categories should be large. The loss function consists of several losses of few-shot learning segmentation network. L_c represents intra-class loss within the memory module. L_s represents inter-class loss within the memory module. λ_c and λ_s are the coefficients of intra-class loss and inter-class loss, respectively as shown in Eq. 5.

$$L = L_f + \lambda_c L_c + \lambda_s L_s \tag{5}$$

where L_f is the loss function of the segmentation network. In this research, we use the average of cross-entropy loss of all space positions, as shown in the Eq. 6. Among them, D_q is the real label chart of the query image, and x , y represents spatial location.

$$L_f = -\sum_x \sum_y D_q(x,y)\log D_q^{'}(x,y) + \left\{\left(1-D_q(x,y)\right)\log\left[1-\mathrm{D}_q^{'}(x,y)\right]\right\} \tag{6}$$

L_c is the intra-class loss of the memory module. The purpose is to encourage the network to query the closest memory item in memory module. To achieve the purpose of reducing the differences in intra-class memory items, as shown in the Eq. 7, the L_2 norm is used as the basis for the loss function.

$$\mathrm{L}\epsilon = \sum_t^T \sum_k^K \|\mathbf{q}_t^k - \mathbf{p}_p\|_2 \tag{7}$$

$\mathbf{p}_p$ is the nearest memory item from $\mathbf{q}_t^k$, that is, p is the index of the recent memory items when the matching probability is the largest, as shown in Eq. 8:

$$p = \underset{m\in M}{argmax}\, w_t^{k,m} \tag{8}$$

However, if only intra-class loss is used, all memory prototypes in the memory module will be too similar. This can cause the module to store only one background class where it cannot remember different background characteristics. This is contrary to the original intention of setting the memory module. If the module cannot remember multiple background characteristics, there is no essential difference between the memory module and the masked average pooling. Therefore, it is required to design an inter -class loss function to enable the memory module to remember multiple types of background characteristics, as shown in 9.

$$\mathrm{L_s} = \sum_t^T \sum_k^K \left[\|\mathbf{q}_t^k - \mathbf{p}_p\|_2 - \|\mathbf{q}_t^k - \mathbf{p}_n\|_2 + \alpha\right]_+ \tag{9}$$

Under the premise of multiple backgrounds of memory modules, similar memory items should be assigned to the same class, thereby reducing the number of similar memory items. This requires to record different background information under the condition of the number of memory items. Since L_c make all memory items be one category, $\mathbf{p}_n$ constituted a loss function. $\mathbf{p}_n$ is the second nearest memory item from $\mathbf{q}_t^k$, as shown in 10.

$$n = \underset{m\in M, m\neq p}{argmax}\, w_t^{k,m} \tag{10}$$

The inter-class loss function encourages $\mathbf{q}_t^k$ to be close to the nearest item, and the distance from the secondary item is far away, thereby achieving the purpose of different categories in the memory item. Under the common action of inter -class losses and losses in the class, a background memory module is able to distinguish with small intra class distance and large inter class distance. When the distance between the two memory items is less than the preset threshold, the inter-class loss becomes effective. On the contrary, the inter-class losses is zero, the category of memory items can be controlled within a certain range.

2.4. Domain adaptive improvement

Current research on few-shot learning semantic segmentation is conducted by network structure which is trained to learn the common information between the support set and query set images. Therefore, while the

support branch guidance the query branch is realized, the differences between different test parts remain large. Thus, the common information learned from the network is less.

We propose the few-shot semantic segmentation of domain adaptive method, and use it to reduce the statistical distribution differences in the support sets and query sets. During the training, reducing the statistical differences of support sets and query sets can enable the network easier to learn the common information between different categories, so as to better guide the query branch. At the same time, in the case of a small amount of data, domain adaptive can better take advantage of the small data sets to learn more guidance information. The goal of domain adaptive is to reduce the difference between the distribution between the support set and the query set. In statistics, the common difference measures are maximum mean discrepancy [37][38] and central moment discrepancy[39]. However, these indicators are limited in the segmentation tasks. Even as the above parameters are minimized, the support set and query set cannot be guaranteed to achieve domain alignment. We further develop our proposed work by incorporating the idea of style migration to perform domain alignment operations without changing the semantic characteristics of the image. This is to ensure that the domain alignment does not have an impact on the segmentation. The method we propose is to operate on the frequency domain of the image. Our studies have shown that the semantics of the image are mainly informed by the phase component of the Fourier transformation. When the amplitude component of the image is exchanged with other image amplitude components, the image can still retain most of the semantic information. Inspired by this, we propose to exchange the low-frequency part of the amplitude components between the support image and the query image, thereby reducing the difference between the distribution of the support image and the query image. The goal is to align the domain between the query set and the support set in each training. The support set and query set pictures have the same size ($H \times W$), and the infrared hot picture is a single channel gray picture, set the query image as Q, and the support image is S. In the case of 1-shot, the support image and query image are used for Fourier transformation. As shown in the Eq. 11, $\mathrm{F}(x)(m,n)$ is the Fourier transformation result. Q passed through Fourier transformation, obtained S_p、Q_p as phase map, and S_a、Q_a are the amplitude pictures.

$$\mathrm{F}(x)(m,n) = \sum_{h,w} \mathrm{Q}(h,w) e^{-j2\pi\left(\frac{h}{H}m + \frac{w}{W}n\right)},\ j^2 = -1 \tag{11}$$

In order to adjust the degree of domain adaptation, we define M_β as a square mask with variable size, the center of the mask is the center of the image (0, 0), as shown in Eq. 12. β is the hyper parameter of the size of the mask, which has nothing to do with the size and resolution of the image.

$$M_\beta(h,w) = 1_{(h,w)\in[-\beta H:\beta H,-\beta W:\beta W]} \tag{12}$$

Fig. 6 shows how the low-frequency part of the query image amplitude component is replaced by the low-frequency component of the support image. This reduces the difference between the distribution of the support image and the query image. The change of the low-frequency part of the picture amplitude component does not affect the semantic segmentation. After changing the low -frequency part of the information, it is equivalent to changing information such as the style, brightness and contrast of the picture, and does not affect the semantic segmentation of the defects.

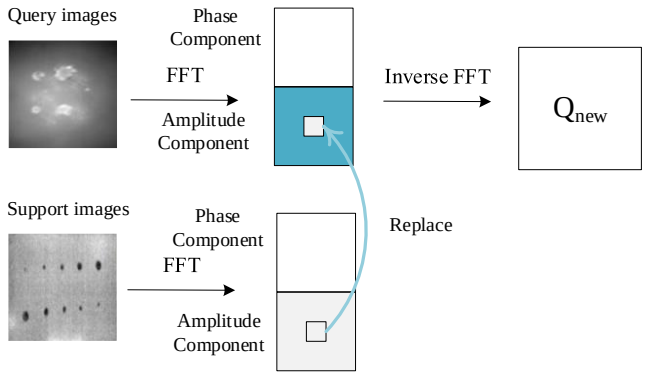


Fig.6 Domain adaptive process schematic diagram

The above process can be represented by Eq. 13. $q \to s$ means that the query image is migrated from the q domain to the s domain; F^{-1} is Fourier inverted transformation; $A(S)$ represents the amplitude component of the support image Fourier transformation; $P(Q)$ is the phase component of the Fourier transformation of the query image.

$$Q_{new}{}^{q \to s} = F^{-1}\left(\left[M_\beta A(S) + (1 - M_\beta) A(Q), P(Q)\right]\right) \tag{13}$$

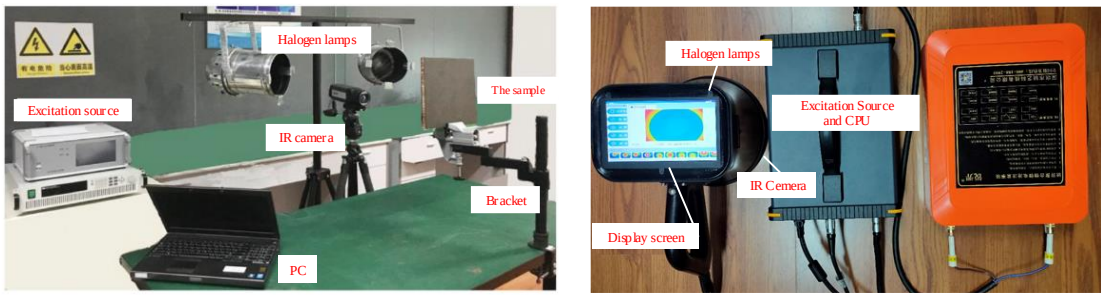
3. EXPERIMENT AND RESULT ANALYSIS

3.1. Experimental Setup and Samples Preparation

Experiments have been completed by two systems. One is a high-precision Optical Pulsed Thermography (OPT) system, which is suitable for laboratory use where it has greater power and irradiation area. The other set is a Portable OPT (POPT) system that is easy to be carried and can facilitates fast detection.

OPT systems is shown in Fig. 7. The infrared thermal camera is FLIRA655SC. The halogen lamp is fixed on the left and right sides of the infrared thermal camera, each side of the halogen light is 1000 W, and the halogen lights are 0.6-1 meters from the test piece. The system controller part consists of a signal generator and adjustable DC power supply. The maximum power of the adjustable DC power supply is 3000 W. The signal generator can generate waveforms such as square waves and sine waves. DC power supply to achieve different incentive methods through signals. At the same time, parameters such as heating cooling time and incentive frequency are set through signal generators.

POPT system combines the controller with the computer, and use the industrial control machine instead of the computer. uses a small infrared thermal camera (MAGNITY MAG-62), and the handheld part contains the source of incentive lights and infrared thermal camera. The incentive source is the six -sided shape of the six halogen lights. The middle part is a infrared thermal camera, and the distance between the lampshade and the test parts at the actual use is 10 cm.



(a) (b)

Fig.7. The experimental systems (a)OPT system (b) POPT system

The proposed model is validated between different test parts to improve the migration capabilities. Due to the training methods of few-shot learning networks, we need to create databases suitable for few-shot learning. Tests are divided into different materials, such as carbon fibre reinforced polymers (CFRP), glass fiber materials, and coating materials. The shape of the test parts is also different. There are common plane types, as well as irregular shapes such as right -angle bend defects. There are also differences in the size and depth of defects. In the source of defects, they can be divided into artificial preparation test parts and natural defects. We divide the different types of test parts into four categories: CFRP test parts, coating test parts, irregular shapes form test parts, rubber and glass fiber material test parts, and divide data sets according to classification. A total of 1284 pictures of various types of test parts were collected, including 323 CFRP test parts, 290 images of coating test parts, 384 images of irregular shapes form test parts, 287 images of rubber and glass fiber material test parts.

3.2. Implementation

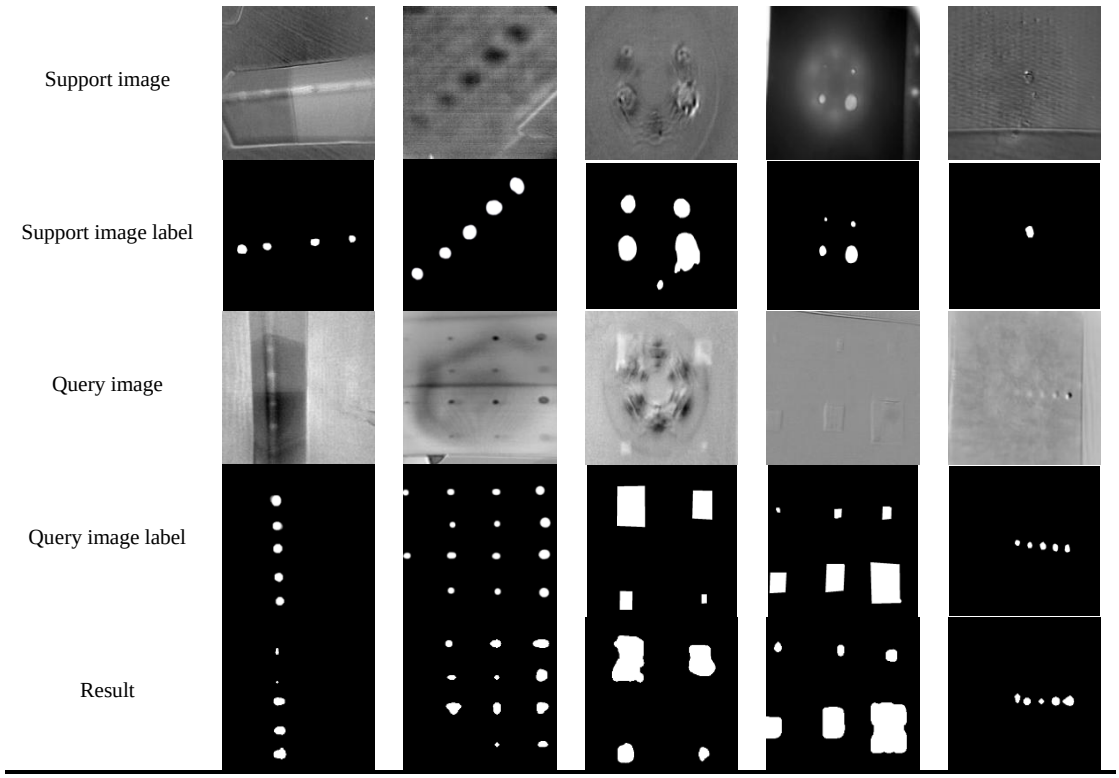
The algorithm model of this article is based on the PyTorch framework. The resolution of the infrared images of OPT system and POPT system is different, with 640 × 480, 288 × 384, respectively. In order to ensure the consistency of the size of the images, the size of the dual-linear interpolation is adjusted to 512 × 512. The experiment was trained using a RTX 2080 GPU, the learning rate was set to 0.0025, and the Batch-size set to 4.

3.3. Results comparison analysis

As shown in Table 1, the 1-shot results of the porposed algorithm on infrared datasets are displayed. The first panel is the support image, the second panel is the support image label, the third line is the query image, that is, the defective images to be detected, the fourth line is the query image label, the fifth line is the result of the algorithm. In the Table 1, the query image number 1 is an irregular shapes test piece, the number 2 is the CFRP test part, the number 3 and 4 are the coating material test parts, and the number 5 is the glass fiber test part.

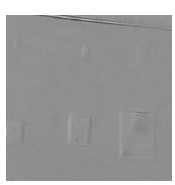
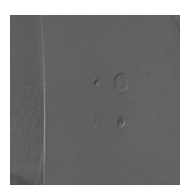
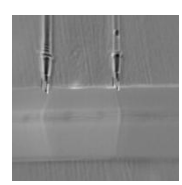
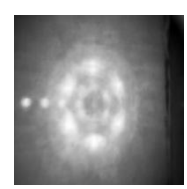
Table 1 Comparison of experimental results

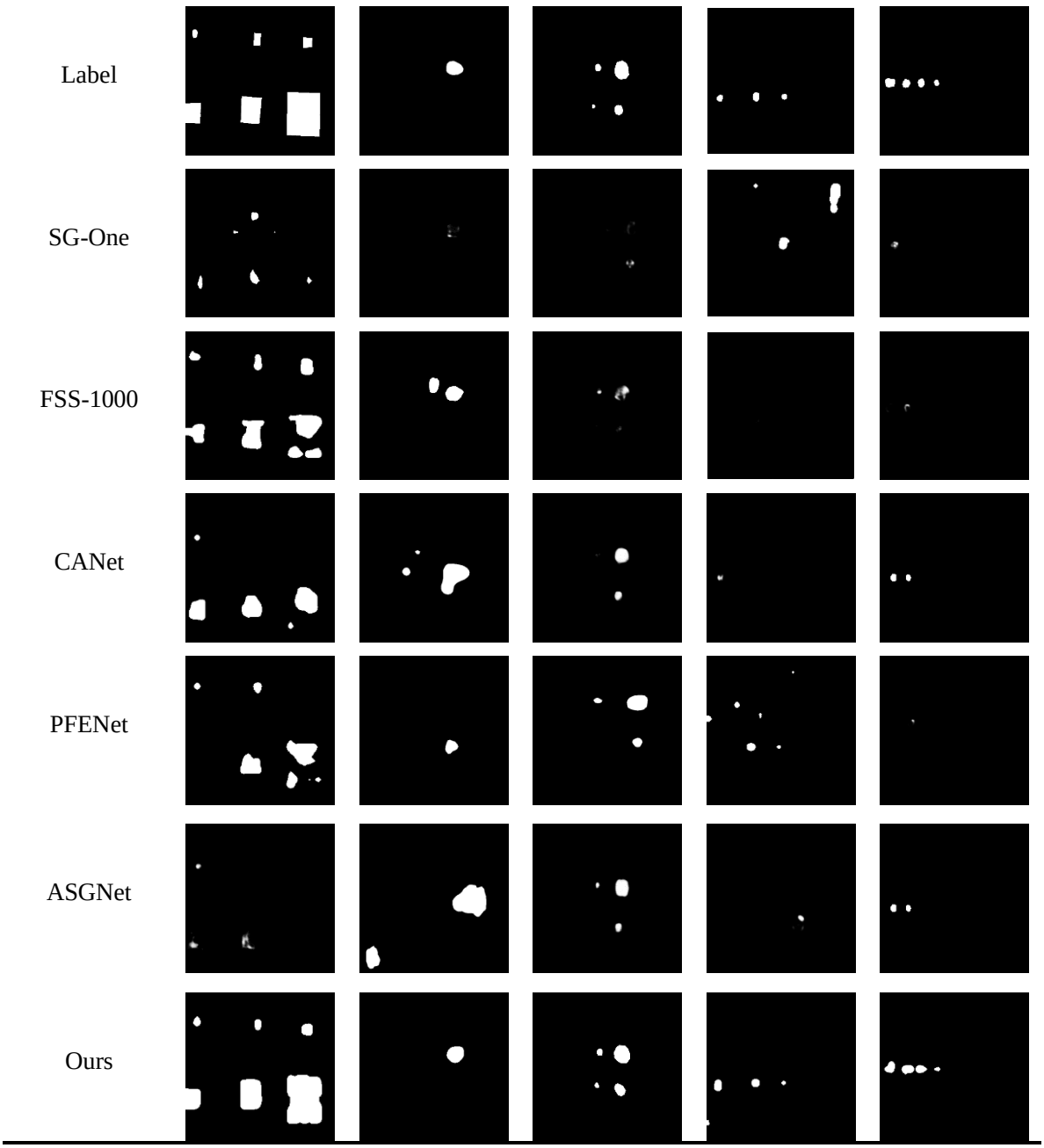
number	1	2	3	4	5



In order to evaluate the results of the proposed algorithm, we compare it with other few-shot learning segmentation networks. The result is shown in Table 2. Among the four types of specimens classified in this paper, the difficulty of CFRP detection is more pronounced and the percentage of difficult samples is not sufficient. Therefore, five difficult samples of complicated shaped specimens are selected to demonstrate the generalization ability of the algorithm and the improvements in defect misses and misdetections. In the 1-shot task, the same specimen support image and support image labels are kept consistent and omitted in the table. The proposed algorithm successfully detects all defects without missing detection. The rest of the algorithms have different degrees of missed or false detections, and the missed detections are more serious than false detections in defect detection. The effectiveness of the proposed algorithm can be visually illustrated by comparing different difficult samples of specimens. Next, we will further examine the performance of the proposed method using objective quantitative analysis.

Table 2 Comparison with other few-shot learning networks

Number	1	2	3	4	5
Test piece picture					



3.4. Quantitative analysis of experimental results

Table 3 shows the quantitative results of the proposed method and the comparison with other few-shot learning segmentation algorithms. From Table 3, it can be seen that the proposed algorithm has significantly outperformed current few-shot semantic segmentation algorithms and is more suitable for IR defect detection tasks.

As can be seen in Table 3, the comparison methods are ineffective except for the proposed method, such as SG-One and FSS-1000. This is because there are thermal diffusion differences in the specimens, so the defect information is easily drowned out by noise and the differences between the defect information and the background information are not easily distinguished and more difficult to be correctly identified.The contrast methods have the difficulty to distinguish the defect and background using only the foreground part of the support image for guidance, and the current few-shot learning segmentation network is mainly designed to segment pictures with more distinct boundaries, such as street scene images, etc., so that the difference between the background information of the images is larger. For defects detection, the difference between defects and non-defects is small. Using the joint guidance of foreground background can effectively increase the metric distance between defects and backgrounds, and effectively segment the defects.

The query image of the test specimen and the support image of the training specimen may differ in material, defect type, and thermal conductivity, and it is difficult to achieve the desired effect with one guidance picture providing guidance information. In contrast, the proposed model can use the foreground information of the guidance picture and the historical background information in the memory module to segment the defects with insufficient guidance information and adapt to a new class more quickly.

Table 3 IOU results for 1-shot setup on IR defect data set

Method	Type 1 specimen	Type 2 specimen	Type 3 specimen	Type 4 specimen	mean
SG-One	54.7	44.9	36.4	40.2	44.05
FSS-1000	58.4	48.7	37.5	43.2	46.95
CANet	59.1	50.3	35.1	41.8	46.58
PFENet	62.1	53.8	33.5	45.6	48.75
ASGNet	64.6	58.4	34.1	46.4	50.88
Ours	**71.4**	**67.6**	**65.8**	**58.9**	**65.93**

Table 4 IOU results for 5-shot setup on IR defect data set

Method	Type 1 specimen	Type 2 specimen	Type 3 specimen	Type 4 specimen	mean	Value Added
SG-One	55.2	47.8	36.8	40.5	45.08	1.03
FSS-1000	64.5	54.8	38.3	47.0	51.13	4.18
CANet	63.0	51.1	40.0	45.1	49.79	3.22
PFENet	66.3	59.2	44.0	47.6	54.26	**5.51**
ASGNet	65.4	60.4	40.2	48.1	53.51	2.64
Ours	**74.6**	**74.0**	**69.7**	**61.3**	**69.90**	3.98

As can be seen from Tables 3 and 4, comparing the current few-shot learning semantic segmentation model, the proposed method has achieved the best performance and also obtained the optimal performance on each type of specimen, and in the case of 1-shot and 5-shot, the IOU of the proposed method have improved by 15.05% and 15.64%, respectively. Among them, there is a large improvement for type 3 and type 4 specimens, where defects are drowned with noise and differ greatly from other types of specimens.

As the IOU evaluation index cannot fully reflect the performance in defect detection, especially the balance between the defect leakage rate and false detection rate. Therefore, this paper also introduces the F-score as an indicator. As shown in Table 5 and Table 6, F-score improvement of 19.75% and 19.78% respectively compared to the existing optimal method.

Table 5 F-score results for 1-shot setup on IR defect data set

Method	Type 1 specimen	Type 2 specimen	Type 3 specimen	Type 4 specimen	mean
SG-One	73.26	63.51	57.35	58.27	63.10
FSS-1000	76.70	68.08	59.42	59.68	65.97
CANet	74.70	66.58	55.55	60.50	64.33
PFENet	77.26	72.27	54.10	61.40	66.26
ASGNet	80.74	74.43	50.22	63.36	67.19
Ours	**94.54**	**90.92**	**81.33**	**80.95**	**86.94**

Table 6 F-score results for 5-shot setup on IR defect data set

Method	Type 1 specimen	Type 2 specimen	Type 3 specimen	Type 4 specimen	mean	Value Added
SG-One	76.63	64.85	58.66	57.62	64.44	1.34
FSS-1000	75.92	70.10	58.12	67.96	68.03	2.06
CANet	77.50	70.90	59.90	62.40	67.68	**3.34**
PFENet	80.75	74.74	57.32	62.82	68.91	2.65
ASGNet	84.42	78.30	51.37	63.45	69.39	2.20
Ours	**95.76**	**94.74**	**83.42**	**82.76**	**89.17**	2.23

3.5. Impaction of Domain Adaptive Module

The proposed domain adaptation method aims to make the new query image closer to the domain of the support image after domain adaptation, so that the support image can better guide the query branch. Several related literature have shown that if the domain adaptation method is used for the feature maps generated by the feature extraction module, poor results are obtained in most cases because this method changes the semantic information of the images.

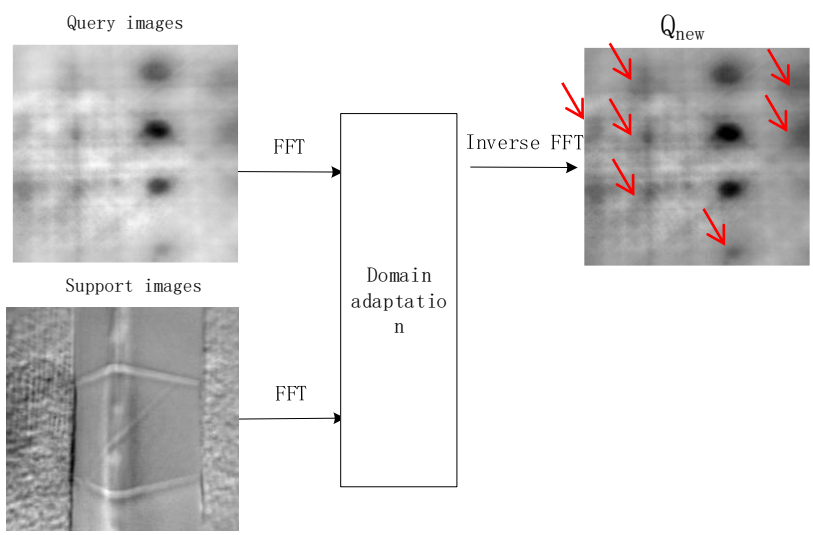
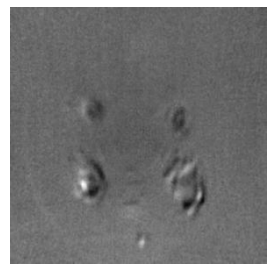


Fig. 8 Visualization of domain adaptive effect

To further illustrate the effectiveness of the domain adaptation module for few-shot learning segmentation networks, this section visualizes the effect of the domain adaptation module in two cases, case 1 can easily segment the support set images with defects and it is difficult to segment the query set images with similar background, and case 2 is the opposite version.

(a)



(b)

Fig. 9 Domain adaptive case one image settings. (a) query image; (b) support image

As shown in Fig.9, Fig. 9 (a) is the query image and (b) is the support image, where the gap between the query image and the support image is large, the background and foreground parts of the query image have a high similarity, which usually fails using the general semantic segmentation algorithm. The support image and the query image have great differences in the background aspect and the foreground aspect, if the support image in Fig. 9 is used to extract information to guide the segmentation of the query image, the network is difficult to extract useful guidance information.

Fig. 10 shows the effect of different β parameters on the domain alignment effect, from (a) to (h) are 0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5. It can be seen that when β is 0.01, the domain adaptation has a little effect. When β is raised to 0.1, domain adaption already has a good effect, and the new query image has a great similarity to the support image, which shows significantly difference from the first query image. In the foreground part of the image, the defects of the new query image are similar to the defects of the support image in terms of color texture. In terms of the background, they also transform from almost no texture to similar to the texture of the support image. However, at β = 0.1, the picture has a blurred condition.

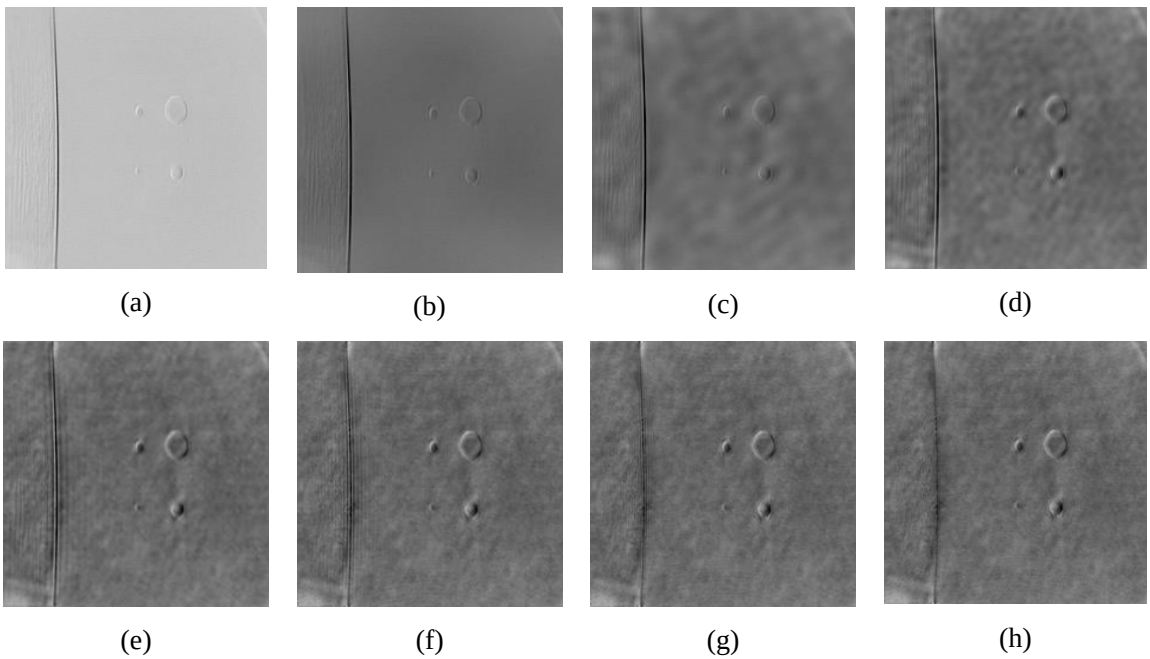
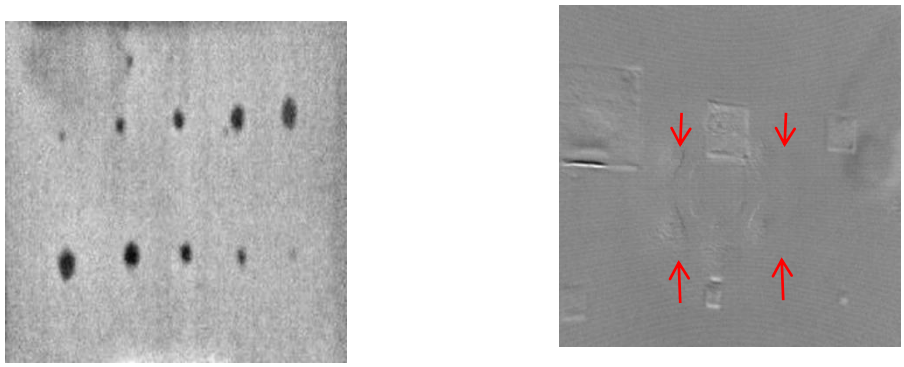


Fig.10 Display of the results of domain adaptation with different β parameters. (a) β =0(Not using domain adaptation);(b) β =0.01;(c) β =0.05;(d) β =0.1;(e) β =0.2;(f) β =0.3;(g) β =0.4;(h) β =0.5

When β = 0.2, the new query image already has good results and can segment the defects using the general semantic segmentation network even without the guidance information provided by few-shot learning. In contrast, the defect information of the original query image differs small from the background, and the difference between the support image query images is too large in the case of the support image. This leads to limited guidance information provided by the support image and possibly wrong guidance information.

In addition to case 1, there exists an opposing situation: it is more difficult for the support image to semantically segment the defects while it is relatively easier for the query image. Few-shot learning aims to learn the commonalities of different types of samples and to be able to better adapt to new samples with a small number of labels when encountering new types of samples. Without domain adaptation for case II, the guidance information provided by the support image of difficult samples is insufficient to be utilized.

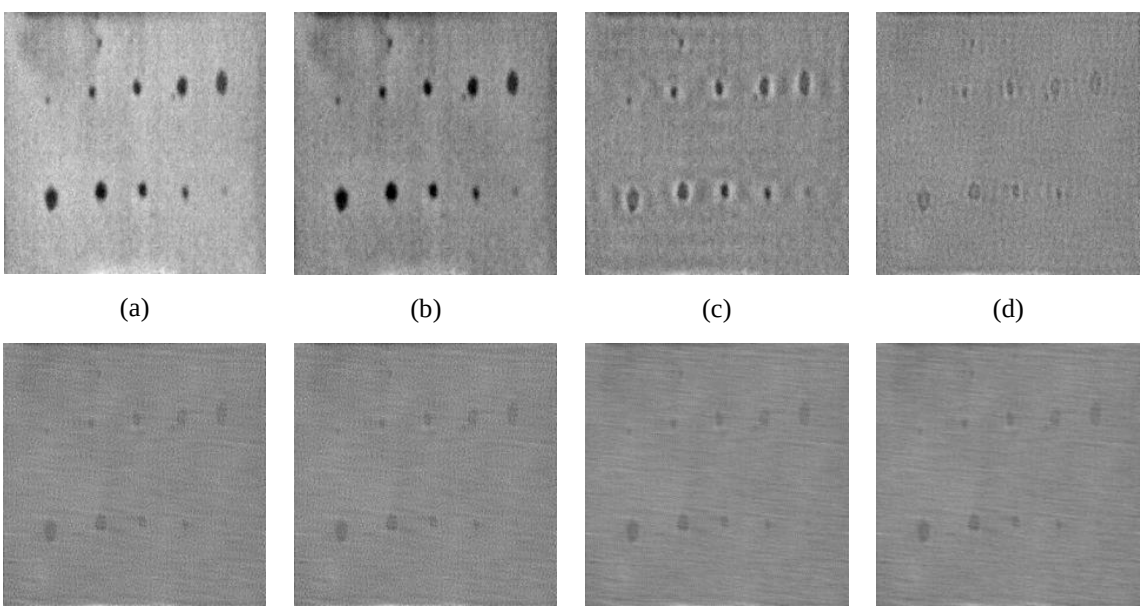
Through the domain adaptive operation, both the support image and the query image become difficult to segment, so that the information of the support image can be better utilized. At the same time, the difficulty of extracting common information by few-shot learning networks decreases, which also improves the network's ability to mine difficult samples. Converting simple samples into difficult samples and increasing difficult samples is beneficial for improving defect detection. Although this seems counter-intuitive, in the field of non-destructive testing difficult samples are easy to miss, which can have serious consequences. In addition, the number of difficult sample data is considerably lesser in the total sample. This is illustrated in Fig. 11. Fig. 11 (a) shows the query image and Fig. 11 (b) shows the support image. The support image is not only difficult for semantic segmentation, but also has lamp shadow interference (as indicated by the red arrow in the figure). This is the reflection caused by the specimen surface being too smooth when the halogen lamp heats the specimen.



(a) (b)

Fig. 11 Domain adaptive case II image settings. (a) query image; (b) support image

Fig. 12 shows the effect of different β parameters on the domain alignment effect, from (a) to (h) are 0, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, we can see that when β is 0.01, the domain adaption has a little effect. Compared with the original image, background color part is more similar and it is still easy for semantic segmentation. When β is raised to 0.1, the picture has some blurring. When β =0.2, the new query image already has a good effect and has a great similarity with the support image.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 12 Display of the results of domain adaptation with different β parameters. (a) β =0(Not using domain adaptation);(b) β =0.01;(c) β =0.05;(d) β =0.1;(e) β =0.2;(f) β =0.3;(g) β =0.4;(h) β =0.5

Since β is ranged from 0.01, 0.05, 0.1, 0.2, the new query image gradually becomes difficult samples. In most few-shot learning network models, the domain difference between the support image and the query image is large, leading to unsatisfactory segmentation results. The domain adaptive approach in this paper narrows the

gap between different domains, which is well targeted for the characteristics of small number of specimens and large differences among specimens in the field of infrared nondestructive testing. The choice of β is crucial. If the parameter is too small, the effect of domain adaptation cannot be played, but if the parameter is set too large, it will lead to artifacts. Thus, β is set to 0.15.

3.6. Impaction of Foreground Background Joint Guidance Module

In order to verify the effectiveness of the storage module, this section incorporates the storage module ablation experiment. To avoid the influence of other modules, the domain adaptation module is removed from the model in this section. In Table 7, foreground indicates that the support branch is guided by foreground only for the query branch, and foreground + background i.e., foreground background jointly guides the query branch.

Table 7 IOU indicators for storage module ablation experiments

Guiding characteristics		Type 1 specimen	Type 2 specimen	Type 3 specimen	Type 4 specimen
1-shot	foreground	61.48	54.14	37.23	45.55
	foreground + background	67.99	63.38	60.95	53.81
5-shot	foreground	59.73	56.24	36.68	45.49
	foreground + background	69.41	67.11	64.49	55.65

As can be seen from the table, the IOU is enhanced using the joint guidance method at both the 1-shot and 5-shot settings. Among them, Type 3 specimens show the most significant enhancement, which is due to the fact that irregularly shaped specimens have a different structure from other types of specimens, have a larger difference in thermal conductivity, and have defects that are highly similar to the background because of thermal noise. Irregularly shaped specimens are more different from other types of specimens, and it is difficult to achieve ideal results on both types of specimens using most semantic segmentation networks for training, and usually a set of parameters or an algorithm is only applicable to one type of specimen. Using only the foreground to guide the query branch, the support graph cannot provide effective guidance information for the query branch, leading to the failure of semantic segmentation.

3.7. Loss function discussion

The key to the ability of the joint guidance module to function is the ability of the background storage module to update the memory prototype with training, requiring a loss function that enables the memory module to remember more types of background information, with each type of information being representative.

The experimental results are shown in Table 8. The storage module cannot be updated properly when the intra-class loss and inter-class loss are removed, which is theoretically similar to the effect of removing the module.

With the inclusion of only intra-class loss functions, the stored prototypes within the storage module tend to be aggregated into one class, with each memory prototype being similar to the others, similar to the treatment where only foreground-guided features are retained. When the interclass loss function is added, the IOU and F-score improve by 4.46% and 6.42%, respectively, indicating that the combined effect of the intraclass and interclass loss functions achieves the effect of storing multi-class background features and a significant improvement in network performance.

Table 8 Comparison of different loss functions at the 1-shot setting

loss functions			Evaluation Indicators	
L_f	L_c	L_s	IOU	F-score
√			50.73	66.25
√	√		51.60	66.84
√		√	55.19	72.67
√	√	√	**60.65**	**78.11**

4. CONCLUSION AND FEATURE WORK

This paper has proposed an end-to-end two-branch few-shot learning segmentation network that uses few-shot learning to solve the problems of insufficient data volume and inadequate generalization performance for multiple types of specimens. By improving the support branch, the foreground background joint guidance is used to better provide guidance information for the query branch and increase the defective and non-defective metric distance to reduce the defect leakage and false detection rate.

The model ignores the temporal dimension information of the infrared thermogram sequence for lightweight computation. Future research can improve the detection capability of the network if the temporal dimension information is utilized for fusion. In addition, the existing model uses optical excitation data but it can be extended to data collected by laser excitation, electromagnetic excitation, etc.

Acknowledgements

The work was supported by Defense Industrial Technology Development Program (Grant No. JSZL2019205C003), National Natural Science Foundation of China (No. 61971093, No. 61527803, No. 61960206010). The work was supported by Science and Technology Department of Sichuan, China (Grant No. 2021YFH0036).

References

[1] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

[2] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," preprint arXiv:1412.7062, 2014. 2, 4

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder decoder with atrous separable convolution for semantic image segmentation," arXiv preprint arXiv:1802.02611, 2018. 2

[4] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

[5] Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1925-1934).

[6] Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2881-2890).

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, 2015, pp. 234–241.

[8] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In CVPR, pages 2881–2890, 2017. 1, 2

[9] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In ECCV, pages 801–818, 2018. 1, 2
[10] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In ICCV, pages 764–773, 2017. 2
[11] Q Xu, B., Chen, H., Mo, Y. L., & Zhou, T. (2018). Dominance of debonding defect of CFST on PZT sensor response considering the meso-scale structure of concrete with multi-scale simulation. Mechanical Systems and Signal Processing, 107, 515-528.
[12] Ding, Y., Jia, M., Miao, Q., & Cao, Y. (2022). A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. Mechanical Systems and Signal Processing, 168, 108616.
[13] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2016). Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. IEEE transactions on pattern analysis and machine intelligence, 39(4), 652-663.
[14] Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. Advances in neural information processing systems, 30.
[15] Kulis, B. (2013). Metric learning: A survey. Foundations and Trends® in Machine Learning, 5(4), 287-364.
[16] Amirreza Shaban, Shray Bansal, Zhen Liu, Irfan Essa, and Byron Boots. One-shot learning for semantic segmentation. In BMVC, 2017
[17] Rakelly, K., Shelhamer, E., Darrell, T., Efros, A., & Levine, S. (2018). Conditional networks for few-shot semantic segmentation.
[18] Zhang, X., Wei, Y., Yang, Y., & Huang, T. S. (2020). Sg-one: Similarity guidance network for one-shot semantic segmentation. IEEE transactions on cybernetics, 50(9), 3855-3865.
[19] Wang, K., Liew, J. H., Zou, Y., Zhou, D., & Feng, J. (2019). Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9197-9206).
[20] Zhang, C., Lin, G., Liu, F., Yao, R., & Shen, C. (2019). Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5217-5226).
[21] Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R., & Jia, J. (2020). Prior guided feature enrichment network for few-shot segmentation. IEEE transactions on pattern analysis and machine intelligence.
[22] Yang, B., Liu, C., Li, B., Jiao, J., & Ye, Q. (2020, August). Prototype mixture models for few-shot semantic segmentation. In European Conference on Computer Vision (pp. 763-778). Springer, Cham.
[23] Sun, Q., Liu, Y., Chua, T. S., & Schiele, B. (2019). Meta-transfer learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 403-412).
[24] Wu, J., Zhao, Z., Sun, C., Yan, R., & Chen, X. (2020). Few-shot transfer learning for intelligent fault diagnosis of machine. Measurement, 166, 108202.
[25] Ruan, L., Gao, B., Wu, S., & Woo, W. L. (2020). DeftectNet: Joint loss structured deep adversarial network for thermography defect detecting system. Neurocomputing, 417, 441-457.
[26] Karataş, M. A., & Gökkaya, H. (2018). A review on machinability of carbon fiber reinforced polymer (CFRP) and glass fiber reinforced polymer (GFRP) composite materials. Defence Technology, 14(4), 318-326.
[27] Zhang, Z., Wang, Y., & Zhang, Z. (2014). Face synthesis from low-resolution near-infrared to high-resolution visual light spectrum based on tensor analysis. Neurocomputing, 140, 146-154.
[28] Tong, Z., Xie, S., Liu, H., Zhang, W., Pei, C., Li, Y., ... & Takagi, T. (2020). An efficient electromagnetic and thermal modelling of eddy current pulsed thermography for quantitative evaluation of blade fatigue cracks in heavy-duty gas turbines. Mechanical Systems and Signal Processing, 142, 106781.
[29] Chen, X., Flynn, P. J., & Bowyer, K. W. (2003, October). PCA-based face recognition in infrared imagery: Baseline and comparative studies. In 2003 IEEE International SOI Conference. Proceedings (Cat. No. 03CH37443) (pp. 127-134). IEEE.
[30] Tong, Z., Xie, S., Chen, H. E., Qiu, J., Cai, W., Pei, C., ... & Takagi, T. (2022). Quantitative mapping of depth profile of fatigue cracks using eddy current pulsed thermography assisted by PCA and 2D wavelet transformation. Mechanical Systems and Signal Processing, 175, 109139.
[31] Qiao, S., Ma, Y., He, Y., Patimisco, P., Sampaolo, A., & Spagnolo, V. (2021). Ppt level carbon monoxide detection based on light-induced thermoelastic spectroscopy exploring custom quartz tuning forks and a mid-infrared QCL. Optics Express, 29(16), 25100-25108.
[32] Moskovchenko, A. I., Vavilov, V. P., Bernegger, R., Maierhofer, C., & Chulkov, A. O. (2020). Detecting delaminations in semitransparent glass fiber composite by using pulsed infrared thermography. Journal of Nondestructive Evaluation, 39(3), 1-10.
[33] Zheng, C. Y., & Li, H. (2013). Small infrared target detection based on harmonic and sparse matrix decomposition. Optical Engineering, 52(6), 066401.
[34] Cen, Y. G., Zhao, R. Z., Cen, L. H., Cui, L. H., Miao, Z. J., & Wei, Z. (2015). Defect inspection for TFT-LCD images based on the low-rank matrix reconstruction. Neurocomputing, 149, 1206-1215.
[35] Medak, D., Posilović, L., Subašić, M., Budimir, M., & Lončarić, S. (2022). DefectDet: A deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images. Neurocomputing, 473, 107-115.
[36] H. Park, J. Noh, and B. Ham, "Learning memory-guided normality for anomaly detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 14372–14381.
[37] Zappulla, F. IEEE SIGNAL PROCESSING SOCIETY. Ann Arbor, 1001, 48109-2110.

[38] Long, M., Cao, Y., Wang, J., & Jordan, M. (2015, June). Learning transferable features with deep adaptation networks. In International conference on machine learning (pp. 97-105). PMLR.
[39] Zellinger, W., Grubinger, T., Lughofer, E., Natschläger, T., & Saminger-Platz, S. (2017). Central moment discrepancy (cmd) for domain-invariant representation learning. arXiv preprint arXiv:1702.08811.
[40] Guaragnella, G., Morelli, D., D'Orazio, T., Galietti, U., Trentadue, B., & Marani, R. (2022, May). Defect detection by a deep learning approach with active IR thermography. In 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT) (Vol. 1, pp. 27-32). IEEE.
[41] Liu, H., Li, W., Yang, L., Deng, K., & Zhao, Y. (2022). Automatic reconstruction of irregular shape defects in pulsed thermography using deep learning neural network. *Neural Computing and Applications*, *34*(24), 21701-21714.