FEW-SHOT DEFECT SEGMENTATION LEVERAGING ABUNDANT DEFECT-FREE TRAINING SAMPLES THROUGH NORMAL BACKGROUND REGULARIZATION AND CROP-AND-PASTE OPERATION

Anonymous ICME submission

ABSTRACT

In industrial quality assessment, it is challenging to conduct automated and accurate defect segmentation under the condition that abundant defect-free images but very limited anomalous images are available. This paper tackles the challenging few-shot defect segmentation task under such condition. We propose two regularization techniques via incorporating abundant defect-free images into the training of an encoder-decoder segmentation network. We first propose a Normal Background Regularization (NBR) loss which is jointly minimized with the segmentation loss, enhancing the encoder network to produce discriminative representations for normal regions. Secondly, we crop/paste defective regions to the randomly selected normal images for data augmentation and propose a weighted binary cross-entropy loss to enhance the training by emphasizing more realistic crop-and-pasted augmented images based on feature-level similarity comparison. Extensive experiments on MVTec AD and MTSD datasets demonstrate the superiority of the proposed method over the competing methods under few-shot settings.

Index Terms— Defect segmentation, Few-shot segmentation, Industrial image inspection

1. INTRODUCTION

Defect segmentation is essential in various industrial inspection tasks relying on localizing and delineating defective regions that visually appear anomalous against the normal ones. In general, it is tedious and time-consuming to manually identify and annotate the defective regions. Moreover, in real-world applications, it is easy to capture defect-free images but challenging to obtain sufficient anomalous images for training [1, 2]. Hence, building up an effective defect segmentation model using very limited number of annotated anomalous images is valuable. Motivated by these demands, this paper tackles the task of few-shot defect segmentation, i.e., defect segmentation with very few annotated anomalous images.

Defect segmentation is related to the task of Anomaly Detection (AD), which aims at classifying an image to be either normal or anomalous. Most works handle AD by exploiting deep neural networks, e.g., Deep Convolution Autoencoder (DCAE) or Generative Adversarial Network (GAN), which are trained using only normal images [3, 4]. Although these AD methods produce good accuracy in classifying low resolution images, they are rarely formulated to localize the defective regions from high-resolution industrial images. Another related task is semantic segmentation, i.e., predicting the semantic class for every pixel in an image. Semantic segmentation is predominantly tackled using fully convolutional networks (FCN) which exploit encoder-decoder architectures to predict masks from images [5–7]. Supervised training of these models typically requires a large number of $\langle \text{image}, \text{mask} \rangle$ pairs for tuning parameters. However, in real-world industrial applications, only very limited anomalous images with pixel-level annotation masks are available. Hence, the supervised training might suffer from overfitting.

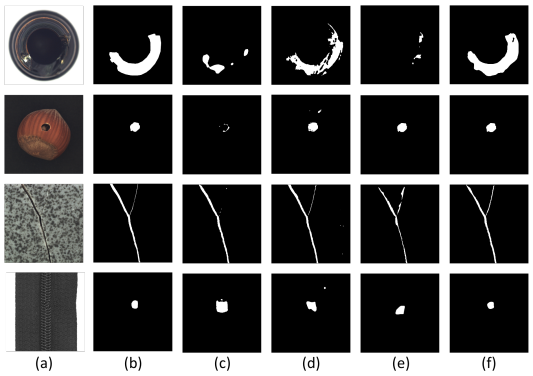


Fig. 1: Illustration of segmentation performance under 1-shot setting: (a) the image; (b) groundtruth; (c) to (f): the predicted masks of (c) Baseline (B) [8]; (d) B+NBR; (e) B+CaP; (f) B+NBR+CaP, respectively.

Recently, few-shot image classification and segmentation [9, 10] have drawn great interest in the computer vision community. The state-of-the-art few-shot approaches typically adopt the meta-learning scheme, which requires an auxiliary dataset with sufficient data to train the meta learner with few-shot learning capability [9, 11, 12]. Differing from the generic few-shot semantic segmentation task, sufficient defect-free training images are available in industrial inspection tasks. In this paper, they are exploited to improve the segmentation performance under few-shot settings.

To address few-shot defect segmentation from high res-

olution images, we propose two regularization methods by leveraging abundant defect-free normal images by involving two input branches to train the segmentation network. One branch is for normal (defect-free) images and the other is for anomalous (defective) images. The proposed methods could be implemented in an encoder-decoder network to improve the defect segmentation performance under few-shot settings. The major contributions are summarized as follows:

1. We propose a novel **Normal Background Regularization (NBR)** to train the encoder network. This loss facilitates the encoder to produce discriminative representations of normal regions by maximizing the similarity between the normal regions within anomalous images and randomly picked normal training images.
2. We propose an effective **Crop-and-Paste (CaP)** operation for data augmentation. Moreover, we develop a weighted binary cross-entropy loss to enhance the influence of more realistic defective images via computing the feature-level similarity between artificially generated and real-captured anomalous images.
3. We adopt a UNet-like encoder-decoder segmentation network backboned by ResNet-34 as the baseline model and demonstrate the effectiveness of the proposed regularization techniques, achieving significantly higher few-shot segmentation accuracy on the MVTec AD and MTSD datasets.

Fig. 1 illustrates the performance of the proposed methods under 1-shot defect segmentation setting. By incorporating NBR and Cap, the segmentation performance is progressively improved compared with the baseline model.

2. METHOD

2.1. Problem Setting

We define our K-shot defect segmentation problem corresponding to C classes of defects and each has K $\langle \text{image}, \text{mask} \rangle$ pairs of training data. Specifically, the training set $\mathcal{D}_{train}$ consists of two subsets: a normal training subset $\mathcal{D}^n_{train} = \{\boldsymbol{I}^n_i\}$ where $i = 1, 2, \ldots, N_n$ and a defective training subset $\mathcal{D}^d_{train} = \{(\boldsymbol{I}^d_{c,k}, \boldsymbol{M}^d_{c,k})\}$ where c ($c = 1, 2, \ldots, C$) and k ($k = 1, 2, \ldots, K$) denote the indexes for the defective categories and the anomalous images, respectively. The testing set $\mathcal{D}_{test}$ contains multiple unseen anomalous images with pixel-wise annotated masks from one of C training defect classes. Both normal and defective training subsets are exploited to train a defect segmentation network whose performance is evaluated on the testing set.

2.2. Baseline Encoder-Decoder Segmentation Network

We adopt an encoder-decoder segmentation network proposed in [8] backboned by ResNet-34 as our baseline model, denoted as **B**. The encoder exploits the convolution blocks of ResNet-34 with fully-connected layers removed. The decoder adopts several decoder blocks to progressively upsample the feature maps. The architecture of the baseline model are specified in the supplementary material.

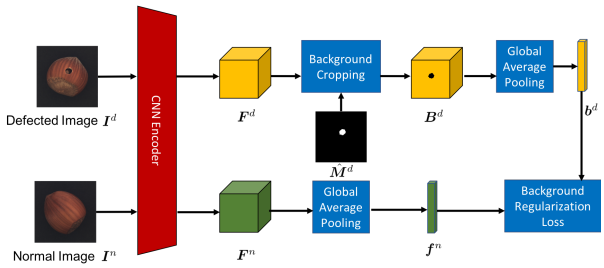


Fig. 2: Illustration of Normal Background Regularization.

2.3. Normal Background Regularization (NBR)

We first propose **N**ormal **B**ackground **R**egularization (NBR) to facilitate the encoder to produce discriminative representations of normal regions by maximizing the similarity between the normal regions within anomalous images and the global regions of normal training images. Specifically, as shown in Fig. 2, assuming batch size is 1, i.e., within one minibatch, we have one normal image $\boldsymbol{I}^n$, one defective image $\boldsymbol{I}^d$ and its defective mask $\boldsymbol{M}^d$. Both $\boldsymbol{I}^n$ and $\boldsymbol{I}^d$ are fed through the encoder to generate feature map $\boldsymbol{F}^n$ and $\boldsymbol{F}^d$, respectively. Then the normal background component $\boldsymbol{B}^d$ is cropped from $\boldsymbol{F}^d$ by elementwisely multiplying $\boldsymbol{F}^d$ with the reversed downsampled mask $\hat{\boldsymbol{M}}^d$:

$$\boldsymbol{B}^d = \boldsymbol{F}^d \odot (\mathbf{1} - \hat{\boldsymbol{M}}^d), \tag{1}$$

where $\hat{\boldsymbol{M}}^d$ is the downsampled version of the original groundtruth mask $\boldsymbol{M}^d$. $\mathbf{1}$ is the mask with the same size as $\hat{\boldsymbol{M}}^d$ filled with ones. $\odot$ denotes elementwise multiplication. Then Global Average Pooling (GAP) is applied on both $\boldsymbol{B}^d$ and $\boldsymbol{F}^n$ to obtain vectorized representations $\boldsymbol{b}^d$ and $\boldsymbol{f}^n$ as:

$$\boldsymbol{b}^d = \text{GAP}(\boldsymbol{B}^d), \quad \boldsymbol{f}^n = \text{GAP}(\boldsymbol{F}^n). \tag{2}$$

To encourage alignment of normal regions, a normal background regularization loss is proposed as the negative cosine similarity between $\boldsymbol{b}^d$ and $\boldsymbol{f}^n$ as:

$$\mathcal{L}_{\text{NBR}} = -\frac{(\boldsymbol{b}^d)^{\text{T}}(\boldsymbol{f}^n)}{||\boldsymbol{b}^d||_2 \cdot ||\boldsymbol{f}^n||_2}. \tag{3}$$

The normal background regularization loss is jointly minimized with the segmentation loss proposed in the subsequent subsection.

2.4. Crop-and-Paste (CaP)

We propose the second regularization method called **C**rop-**a**nd-**P**aste (CaP) operation. As shown in Fig. 3, the defective

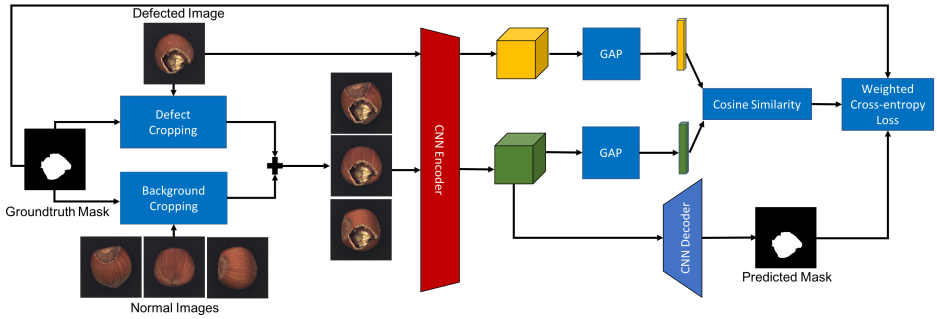


Fig. 3: The illustration of Crop-and-Paste (CaP) operation.

region in an anomalous image $\boldsymbol{I}^d$ is cropped out and pasted upon a normal image $\boldsymbol{I}^n$. Such operation is conducted using the groundtruth mask $\boldsymbol{M}^d$ by:

$$\boldsymbol{I}^d_{\text{CaP}} = \boldsymbol{I}^d \odot \boldsymbol{M}^d + \boldsymbol{I}^n \odot (\boldsymbol{1} - \boldsymbol{M}^d). \tag{4}$$

The resulted augmented image $\boldsymbol{I}^d_{\text{CaP}}$ is fed through the encoder and the decoder networks to generate predicted defect mask $\tilde{\boldsymbol{M}}^d_{\text{CaP}} \in [0, 1]$ using sigmoid activation function on the final output layer. We propose a weighted binary cross-entropy loss by pixelwisely comparing the groundtruth mask $\boldsymbol{M}^d$ and $\tilde{\boldsymbol{M}}^d_{\text{CaP}}$:

$$\begin{aligned}\mathcal{L}_{\text{WBCE}} = -\lambda \sum_{(w,h)} \Big[\boldsymbol{M}^d(w,h) log\Big(\tilde{\boldsymbol{M}}^d_{\text{Cap}}(w,h)\Big) + \\ \Big(1 - \boldsymbol{M}^d(w,h)\Big) log\Big(1 - \tilde{\boldsymbol{M}}^d_{\text{CaP}}(w,h)\Big)\Big],\end{aligned} \tag{5}$$

where w and h denote the width and height coordinate, respectively. The weight λ is defined as the cosine similarity between two vectors after global average pooling $\boldsymbol{F}^d$ and $\boldsymbol{F}^d_{\text{DCP}}$:

$$\lambda = \frac{|\text{GAP}(\boldsymbol{F}^d)^{\text{T}} \text{GAP}(\boldsymbol{F}^d_{\text{CaP}})|}{||\text{GAP}(\boldsymbol{F}^d)||_2 \cdot ||\text{GAP}(\boldsymbol{F}^d_{\text{CaP}})||_2}. \tag{6}$$

λ quantifies the contribution of $\mathcal{L}_{\text{WBCE}}$ by the similarity between $\boldsymbol{F}^d$ and $\boldsymbol{F}^d_{\text{CaP}}$ after the CaP operation. λ is calculated to quantify the "realistic degree" of the augmented image by comparing it to the original anomalous image. The augmented image is more realistic if it is highly similar to the original anomalous image. Hence, a bigger λ reflects a relatively realistic augmented anomalous image and its loss contributes more to tune the network parameters.

For each minibatch iteration, we set a probability parameter which is uniformly drawn from the range of [0, 1]. If the probability parameter exceeds a fixed threshold, the CaP operation is applied. Otherwise, we feed the original feature map of anomalous image $\boldsymbol{F}^d$ through the decoder to obtain the predicted mask $\tilde{\boldsymbol{M}}^d$ and then calculate the standard binary cross-entropy loss as:

$$\begin{aligned}\mathcal{L}_{\text{BCE}} = -\sum_{(w,h)} \Big[\boldsymbol{M}^d(w,h) log\Big(\tilde{\boldsymbol{M}}^d(w,h)\Big) + \\ \Big(1 - \boldsymbol{M}^d(w,h)\Big) log\Big(1 - \tilde{\boldsymbol{M}}^d(w,h)\Big)\Big].\end{aligned} \tag{7}$$

Algorithm 1 Training procedure

Input: Training samples in $\mathcal{D}^n_{train}$ and $\mathcal{D}^d_{train}$.
for number of iterations on defect batch sampling **do**
- Sample a minibatch of $\langle$image, mask$\rangle$ pairs $\{(\boldsymbol{I}^d_1, \boldsymbol{M}^d_1), \ldots, (\boldsymbol{I}^d_m, \boldsymbol{M}^d_m)\}$ from $\mathcal{D}^d_{train}$.

for number of iterations on normal batch sampling **do**
- Sample a minibatch of normal images $\{\boldsymbol{I}^n_1, \ldots, \boldsymbol{I}^n_m\}$ from $\mathcal{D}^n_{train}$.

Normal Background Regularization
- Calculate $\mathcal{L}_{\text{NBR}}(\boldsymbol{I}^n_i, \boldsymbol{I}^d_i)$ using Eqs. (1), (2) and (3)
- Calculate $\mathcal{L}^m_{\text{NBR}} = \sum_{i=1}^m \mathcal{L}_{\text{NBR}}(\boldsymbol{I}^n_i, \boldsymbol{I}^d_i)$.

Crop-and-Paste Operation
if probability $> 50\%$: **then**
- Calculate $\mathcal{L}_{\text{WBCE}}(\boldsymbol{I}^n_i, \boldsymbol{I}^d_i, \boldsymbol{M}^d_i)$ using Eqs. (4), (5) and (6)
- Calculate $\mathcal{L}^m_{\text{WBCE}} = \sum_{i=1}^m \mathcal{L}_{\text{WBCE}}(\boldsymbol{I}^n_i, \boldsymbol{I}^d_i, \boldsymbol{M}^d_i)$.
- Calculate the overall loss of the minibatch: $\mathcal{L} = \mathcal{L}^m_{\text{NBR}} + \mathcal{L}^m_{\text{WBCE}}$.

else
- Calculate the original binary cross-entropy loss $\mathcal{L}_{\text{BCE}}(\boldsymbol{I}^d_i, \boldsymbol{M}^d_i)$ using Eq. (7).
- Calculate $\mathcal{L}^m_{\text{BCE}} = \sum_{i=1}^m \mathcal{L}_{\text{BCE}}(\boldsymbol{I}^n_i, \boldsymbol{I}^d_i, \boldsymbol{M}^d_i)$.
- Calculate the overall loss of the minibatch: $\mathcal{L} = \mathcal{L}^m_{\text{NBR}} + \mathcal{L}^m_{\text{BCE}}$.

end if
- Tune the network by descending the gradients of $\mathcal{L}$.

end for
end for

To combine both NBR and CaP, we minimize the overall loss $\mathcal{L}$ defined as the summation of $\mathcal{L}_{\text{NBR}}$ and the segmentation loss depending on the probability parameter. Algorithm 1 summarizes the overall training procedure.

3. EXPERIMENTS

3.1. Dataset

MVTec Anomaly Detection dataset (MVTec AD) [1]. MVTec AD contains 5354 high-resolution images of multiple object or texture categories. For each category, it contains normal images and anomalous images. Overall, MVTec AD contains over 70 different types of defects such as holes, contaminations, scratches, and etc.

Magnetic Tile Surface Defect dataset (MTSD) [2]. MTSD contains 1344 magnetic tile surface images which are sepa-

rated into six subsets of different defect types: "Free" (Normal), "Blowhole", "Crack", "Fray", "Break" and "Uneven".

3.2. Experiment Settings

For MVTec AD dataset, we first conduct the experiment on few-shot defect segmentation. For K-shot defect segmentation setting, K $\langle \mathrm{image}, \mathrm{mask} \rangle$ pairs are randomly selected for each defect type to construct the defect training subset and the normal training subset is constructed using the defect-free training images. The remaining anomalous images are used for performance evaluation. The second experiment is on few-shot anomaly detection. Since the proposed method is formulated for defect segmentation, a testing image is predicted as anomalous if at least one of its pixels is predicted as defective. The anomaly score for the image is defined as the area of the defective regions. For MTSD dataset, we conduct 5, 10 and 15-shot defect segmentation experiment.

All the images are resized to 512×512. The batch size is set as 2 for 1-shot and 4 for more than 1-shot settings, respectively. Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ are exploited and the learning rate is set as 0.0001. The number of training iterations is set as 1350. For our proposed method, the probability of adopting the crop-and-paste is set as 50%. To build strong benchmarking methods, the CaP is also applied for the benchmarking methods but they are trained using the standard BCE loss. Random horizontal and vertical flip are conducted for data augmentation.

3.3. Performance Metrics

The Intersection-Over-Union (IOU) and Dice Coefficient (DC) are adopted to evaluate defect segmentation segmentation performance:

$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{TP+FN+FP}},\ \mathrm{DC} = \frac{2\mathrm{TP}}{2\mathrm{TP+FN+FP}}. \tag{8}$$

where TP, FP, and FN denotes the number of pixels in the predicted mask which are true positive, false positive and false negative compared with the groundtruth mask, respectively. For few-shot anomaly detection, the average classification accuracy and the area under the ROC curves are adopted as the performance metrics.

3.4. Experiments on MVTec AD dataset

Comparison on Few-Shot Defect Segmentation. Table 1 shows the mean IOU and DC over MVTec AD categories. It is observed that the proposed method significantly outperforms all the benchmarking methods under both few-shot settings. It is also noted that even when more annotated training data is provided, our method could still improve the segmentation performance. From the experimental results, the proposed methods effectively regularizes a high capacity baseline model **B** and significantly boost the segmentation performance. Fig. 4 shows the qualitative segmentation results of compared methods under 1-shot setting. Providing only a single annotated training image, our model could produce better generalization segmentation performance than all the benchmarking methods, especially on the fine details such as object edges and thin cracks. The specific performance for each category of MVTec AD dataset is provided in the supplementary material.

Table 1: Mean IOU and DC for experiments on MVTec AD

Method	1 shot		5 shot	
	IOU	DC	IOU	DC
TernausNet-11 [13]	0.3137	0.4949	0.4083	0.6023
TernausNet-16 [14]	0.2917	0.4877	0.3839	0.5950
U-Net [5]	0.3136	0.4989	0.4049	0.6033
Attention U-Net [7]	0.2824	0.4730	0.3701	0.5827
U-Net++ [6]	0.3458	0.5463	0.4377	0.6482
AlbuNet-34 (B) [8]	0.3327	0.5078	0.4217	0.6110
B+NBR+CaP (ours)	**0.4919**	**0.6296**	**0.5916**	**0.7320**

Comparison on Few-Shot Anomaly Detection. Table 2 records the average classification accuracy and the areas under the ROC curves, respectively. The ROC curves are provided in the supplementary material. It is observed that our model produces higher classification accuracy than all the benchmarking methods. Under 1-shot setting, the proposed model produces 88.56% and 92.19% in classification accuracy and AUC, respectively. Under 5-shot setting, we reach 94.12% and 97.03%. We achieve such high classification accuracy at very low annotation cost by leveraging sufficient normal defect-free training data.

Comparison with unsupervised AD methods. We also compare our method with several unsupervised AD methods in terms of average AUC over the categories as shown in Table 3. The average AUCs of SSIM and L2 Autoencoder, AnoGan, and CNN Feature Dictionary (CFD) are calculated using the results reported in [1] and that of DeepSVDD [4] is reported in [15]. Our method outperforms the best benchmarking method by around 5% and 10% in AUC under the settings of 1 shot and 5 shot, respectively.

3.5. Experiments on MTSD dataset

In this subsection, we conduct the experiment using MTSD dataset. Table 4 shows the mean IOU and DC under 5-shot, 10-shot and 15-shot settings, respectively. It is observed that our method can generally produce better segmentation performance than all the related methods under the few-shot settings. The qualitative results on MTSD dataset can be viewed in the supplementary material.

3.6. Ablation Study and Discussion

In this subsection, some ablation studies and discussions are conducted using MVTec AD dataset. We mainly con-

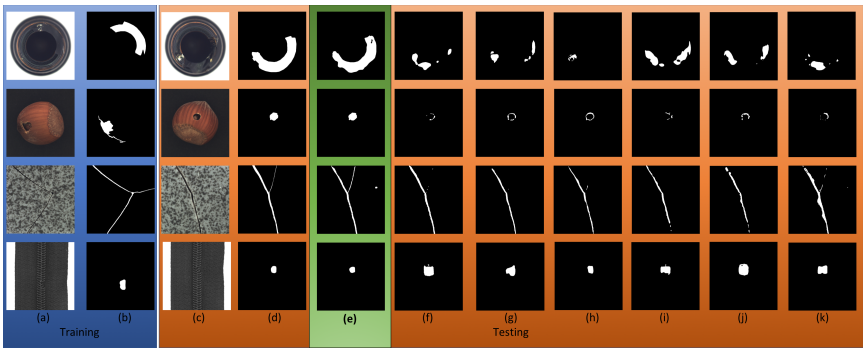


Fig. 4: The qualitative segmentation results under 1-shot defect segmentation setting. (a) and (b): the training image and its groundtruth; (c) and (d): the testing image and its groundtruth; (e) to (i): the predicted masks of (e) our method (B+NBR+CaP); (f) B [8]; (g) TernausNet-11 [13]; (h) TernausNet-16 [14]; (i) U-Net [5]; (j) Attention U-Net [7]; (k) U-Net++ [6], respectively.

Table 2: Classification accuracy (ACC) and area under ROC curves (AUC) under 1-shot and 5-shot settings.

Method	1 shot		5 shot	
	ACC	AUC	ACC	AUC
TernausNet-11 [13]	78.57%	83.13%	90.51%	94.11%
TernausNet-16 [14]	76.99%	81.83%	89.04%	93.58%
U-Net [5]	80.26%	86.25%	90.88%	93.98%
Attention U-Net [7]	74.94%	81.22%	86.69%	88.84%
U-Net++ [6]	81.23%	87.10%	90.01%	94.47%
AlbuNet-34 (B) [8]	77.78%	83.97%	85.96%	89.47%
B+NBR+CaP (ours)	**88.56%**	**92.19%**	**94.12%**	**97.03%**

Table 3: Comparison with unsupervised AD methods.

AE(SSIM)	AE(L2)	AnoGAN	CFD	Deep SVDD	Ours (1 shot)	Ours (5 shot)
87%	82%	74%	78%	72%	**92.19%**	**97.03%**

sider four ablation models: **B**-the baseline model; **B+NBR**-the baseline model with normal background regularization; **B+CaP**- the baseline model with only crop-and-paste operation; **B+NBR+CaP**-the baseline model with both normal background regularization and crop-and-paste operation.

Table 5 shows the mean IOU and DC for all the ablation models, respectively. It is clearly observed that **B+NBR**, **B+CaP** and **B+NBR+CaP** significantly outperform the original baseline model **B** by big margin. Secondly, **B+NBR+CaP** further boosts the performance of **B+NBR** by around 7% and 4% and that of **B+CaP** by around 5% and 1% under 1-shot and 5-shot settings, respectively. This indicates that the proposed regularization methods could jointly improve the performance of the baseline model. To observe the regularization performance, the curve of mean IOUs is plotted using the testing set of category "Carpet" during the training of the four ablation models. Fig. 5a and 5b show the mean IOUs under 1-shot and 5-shot settings, respectively. Compared with **B**, incorporating either **NBR** or **CaP** not only achieves better generalization performance, but also accelerates the training convergence process by adding effective regularization.

We further conduct two ablation experiments on the effect of the parameter λ (Eq. (6)) and on the probability of adopting CaP. From the experiments, it is noted that by introducing the weight λ, the segmentation performance is improved for both 1-shot and 5-shot settings. It is also observed that the performance of the proposed method generally remains robust to changes of probability values and the best performance can be achieved with the probability set as 50%. The specific results are in the supplementary material.

Table 4: Mean IOU and DC for experiments on MTSD

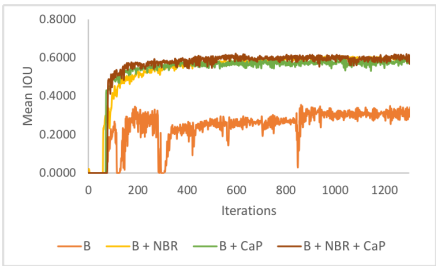
Method	5 shot		10 shot		15 shot	
	IOU	DC	IOU	DC	IOU	DC
TernausNet-11 [13]	0.0852	0.1264	0.1501	0.2131	0.2479	0.3429
TernausNet-16 [14]	0.0976	0.1410	0.1287	0.1864	0.1900	0.2692
U-Net [5]	0.0774	0.1102	0.1279	0.1827	0.2175	0.2973
Attention U-Net [7]	0.0561	0.0814	0.1032	0.1471	0.2322	0.3192
U-Net++ [6]	0.1114	0.1609	0.1338	0.1900	0.2263	0.3220
B (AlbuNet-34) [8]	0.1305	0.1847	0.1701	0.2332	0.2914	0.3861
B+NBR+CaP (ours)	**0.2555**	**0.3459**	**0.3414**	**0.4437**	**0.4084**	**0.5166**

Table 5: Mean IOU and DC for ablation study.

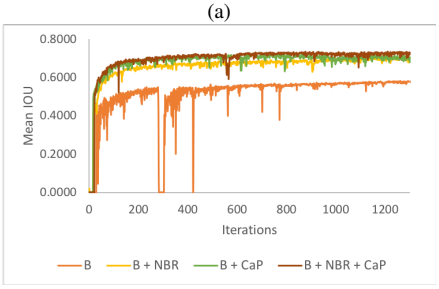
Method	1 shot		5 shot	
	IOU	DC	IOU	DC
B	0.3190	0.3936	0.5507	0.6537
B+NBR	0.4241	0.5245	0.6039	0.7101
B+CaP	0.4427	0.5413	0.6314	0.7356
B+NBR+CaP	**0.4919**	**0.5944**	**0.6445**	**0.7479**

4. CONCLUSION

Towards few-shot defect segmentation, we have proposed two effective regularization methods, namely Normal Background Regularization (NBR) and Crop-and-Paste (CaP) operation to train an encoder-decoder segmentation network. Both methods effectively exploit the abundant defect-free normal images to alleviate model overfitting due to very limited annotated anomalous training images. NBR facilitates the encoder network to produce discriminative representations for normal



(a)



(b)

Fig. 5: The mean IOUs using the testing set of category "Carpet" under (a) 1-shot setting and (b) 5-shot setting.

regions. CaP exploits normal training images for data augmentation and enhance the training of the segmentation networks by emphasizing more realistic augmented images. Extensive experiments are conducted on two industrious datasets and show the superiority of the proposed method over multiple benchmarking methods under the settings of few-shot defect segmentation.

5. REFERENCES

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger, "MVTec AD–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9592–9600.

[2] Yibin Huang, Congying Qiu, and Kui Yuan, "Surface defect saliency of magnetic tile," *The Visual Computer*, vol. 36, no. 1, pp. 85–96, 2020.

[3] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2898–2906.

[4] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft, "Deep one-class classification," in *International conference on machine learning*, 2018, pp. 4393–4402.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[6] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, 2019.

[7] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[8] Alexander Buslaev, Selim S Seferbekov, Vladimir Iglovikov, and Alexey Shvets, "Fully convolutional network for automatic road extraction from satellite imagery.," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 207–210.

[9] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al., "Matching networks for one shot learning," in *Advances in neural information processing systems*, 2016, pp. 3630–3638.

[10] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng, "PANet: Few-shot image semantic segmentation with prototype alignment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9197–9206.

[11] Spyros Gidaris and Nikos Komodakis, "Dynamic few-shot visual learning without forgetting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4367–4375.

[12] Chi Zhang, Guosheng Lin, Fayao Liu, Rui Yao, and Chunhua Shen, "CANet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5217–5226.

[13] Vladimir Iglovikov and Alexey Shvets, "Ternausnet: U-Net with VGG11 encoder pre-trained on imagenet for image segmentation," *arXiv preprint arXiv:1801.05746*, 2018.

[14] Alexey A Shvets, Alexander Rakhlin, Alexandr A Kalinin, and Vladimir I Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 624–628.

[15] Kang Zhou, Yuting Xiao, Jianlong Yang, Jun Cheng, Wen Liu, Weixin Luo, Zaiwang Gu, Jiang Liu, and Shenghua Gao, "Encoding structure-texture relation with P-Net for anomaly detection in retinal images," *arXiv preprint arXiv:2008.03632*, 2020.