DCSNet: A Surface Defect Classification and Segmentation Model by One-Class Learning

Wang Yuxiang 1,2, Mu Shiyi 1,2 Xie Xiang 1,3 and Huang Shanshan 4

1. School of Information and Electronics, Beijing Institute of Technology, Beijing, China.
2. Shanghai University, Shanghai, China.
3. Shenzhen Research Institute, Beijing Institute of Technology, Shenzhen, China.
4. Xi’an Technological University, Xi’an, China.

Email: wyx_bit@163.com, mushiyi@shu.edu.cn, xiexiang@bit.edu.cn, mochengyv@163.com

Abstract. Researches in surface defect classification and segmentation technology have been seen significant progress in recent years. However, there are few works on One-Class learning in this direction by a single model. In previous researches, some problems remain unsolved in the surface defect detection methods, e.g. the training needs a large number of samples and these models cannot classify and locate the surface defect accurately, etc. The main contribution in this work is that we summarize the overall ideas of previous research in network design and propose a multi-task model which could be trained only using a few of positive samples. Meanwhile, the experiments on AITEX detection datasets[1] which get 84.4% DR, 4.4% FAR and 34.2% MIOU, and conduct an ablation experiment in real industrial product dataset to validate the effect of different backbones on DCSNet. It’s worth mentioning that DCSNet provides a solution to the task of surface defect classification and segmentation based on One-Class learning. The code will be open source in https://agit.ai/wyxxy/zhengtu.

1. Introduction
In recent years, as deep learning has shown its strongly feature extraction abilities. Neural networks have achieved extremely high accuracy in supervised surface defect detection. Liang et al.[2] proposed a ShuffleNet v2-based classifier to classify the inkjet code defects of plastic containers in a complex background. Yu et al.[3] used a two-stage FCN network and achieved an average accuracy of 95.99% on the DAGM2007 surface defect dataset. Ding et al.[4] designed a detection network for PCB surface defects and achieved 98.9% mAP by introducing multi-scale feature pyramids and online hard negative mining.

The above application scenarios are all based on supervised learning. However, in real industrial scenes, supervised defect detection faces the following three problems: lack of sufficient negative samples, manual Labelling is Expensive and poor model generalization ability[5].

Overview of the proposed approach:
Supervised learning is constrained by the problems above and is difficult to implement in actual production scenarios. In the field of semi-supervised and unsupervised anomaly detection, there have been a series of practical applications in recent years [6-7]. However, there are very few works in the...
direction of unsupervised surface defect classification and segmentation [8]. In order to solve the problem of supervised learning in specific scenarios, this paper propose a multi-task benchmark model based on One-Class Learning, which only uses a few of positive samples to train model and realize the classification and segmentation of surface defects in industrial environment.

![Image](image.png)

**Figure 1.** The first row is the pictures of different kinds of products, the second row is the result of the image repair module, and the third row is the result of the segmentation module. The data shown are from references[9]

2. Related Work

2.1. Image Inpainting
Initially proposed in [10], the theory behind GAN is based on a competition of two networks within a zero-sum game framework. GANs effectively minimize a Jensen-Shannon divergence, thus generating in-distribution images[11]. In the research of image inpainting, the commonly used method is to use the Context-Encoder to predict the content of the hole that is deducted from the features extracted by the convolutional network. Yu et al. [12] proposed a context-based attention layer to search for the background patch set with the highest speed in rough prediction. Yan et al. [13] proposed shift-net by introducing shift operation and guidance loss as an entry point to improve the quality of the image restoration model. Liu et al. [14] introduces a coherent semantic attention layer to integrate the context of the image hole location and accurately predict the connection between the missing semantic feature part and the context.

2.2. Anomaly Detection
Anomaly detection is an important field of machine learning, and it has a wide range of applications in fields such as biomedicine [15-16], video surveillance [17] and public security inspection [18]. Zenati et al.[19] introduced the use of BiGAN[20] in anomaly detection tasks, and established a mapping from image space to latent space and a mapping from latent space to image space in joint training. Akay et al.[21] proposed a new Auto-Encoder training mode, which uses the encoder acts as a discriminator to accelerate network convergence in training phase and use the encoder as a metric network to measure the distance between repair image and input image. Afterwards, Akacy et al. improved the structure of the literature [22], using a jump connection structure, and achieved state-of-the-art effects in the CIFAR-10 dataset, UBA dataset and FFDB dataset [18]. Xia, Zhang et al. [23] proposed a two-stage anomaly detection network. In the first stage, a synthesized image is first generated according to a given segmentation layout map. In the second stage, the Siamese network is used to segment the abnormal area. This method achieves 6% AUPR-Error in CitySpaces[24] and 20% AUPR in StreetHazards[25].

3. Proposed Approach: DSCNet
Our surface defect classification and segmentation system is called DSCNet, the framework is shown in Figure 2. The model as a whole consist of three modules: Image Repair Module, Defect Segmentation Module and Defect Classification Module.

3.1. Image Repair Module

The overall network framework is shown in Figure 1, inspired by image inpainting works[12-14], the part of image repair module is an Auto-Encoder based on generative adversarial structure to construct a mapping between latent positive sample space and image space.

In the training stage, \( X \) is a set of images randomly selected from the training set, and \( C(\tilde{z}|x) \) is random defects artificially synthesized, randomly posted on a random number of images in \( X \). Among them, EN and DE constitute an Auto-Encoder structure, and the D is the discriminator used to identify the quality of the repaired image, thereby improving the generation quality and convergence speed of the entire image repair module[10].

3.2. Defect Segmentation Module

In order to guide the segmentation network and accelerate the convergence speed, the method proposed the aggregation vector \( X_{\text{const}} \) which concatenate the input image \( X \), repaired image \( X_R \) and the difference image \( X_{\text{abs}} \).
Put the modified aggregation vector input the defect segmentation module and get a defect segmentation map. Among them, the structure of the defect segmentation module is shown in Figure 3.

Assuming that the number of down sampling layers is \( N \), the nth up sampling feature map \( S_{up}^n \) is amplified by a transposed convolution, and then added with the \( N - n \) th down sampling feature map. Due to the random shape and size of industrial product defects, the quality of segmented image can be improved by properly expanding the receptive field of convolution [26]. Therefore, after sampling on the image, the proposed model use a self-attention module to further expand the receptive field. After a self-attention module, the \( N + 1 \) th up sampling feature map is obtained. After five up sampling and five down sampling, the defect segmentation map \( X_D \) is output.

3.3. Defect Classification Module

The traditional surface defect detection network usually uses the threshold method to determine whether the input image is a defective image or not[21,22,27]. Inspired by [15], DCSNet is combined with a defect classification module by integrating the features extracted from the upper sampling of the segmentation module and combining the features obtained by its own down sampling convolution.

As shown in Figure 4, the classifier is a down sampling full convolution network with the VGG structure. The up-sampling feature size of the segmenter is consistent with the down-sampling feature map corresponding to the classifier, and the feature map of the classifier corresponding to each layer is spliced with the feature map of the segmenter to form a layer-by-layer feature splicing classifier network. The last layer uses global average pooling to form a two-classification network to determine whether the input image is a defective image.

3.4. Loss Function

Contextual Loss: In the training phase, in order to enable the Auto-Encoder of the defect repair module to fit the positive latent sample space, the repair sample \( X_R \) should be basically equal to the positive sample \( X \). Therefore, referring to the work in pix2pix[28] and use the L1 distance as the evaluation of the similarity between the repaired sample and the positive sample, the loss function is defined as followed:

\[
L_{CON} = \mathbb{E}_{x \sim p_{data}(x)} \left[ \| x - G(\hat{x}) \|_1 \right] \tag{2}
\]

GAN Loss: In case of complex image details, if only contextual loss as the loss function may not generate image details accurately[10]. Therefore, DCSNet use the structure of generating confrontation to accelerate the network fitting the latent positive space.

\[
L_{GAN} = \mathbb{E}_{x \sim p_{data}(x)} \left[ \log D(x) + \log(1 - D(G(\hat{x}))) \right] \tag{3}
\]

Classification Loss: In this paper, the cross entropy is used as the loss function of defect classification module. In the image repair stage, if the artificial defect \( \hat{X}_D \) is generated and pasted on the positive sample image, the sum of the artificially generated defect map \( \sum \hat{X}_D > 0 \). So, the classification label \( \hat{y} \) is defined as follows.

\[
\hat{y} = \begin{cases} 
0 , \quad \sum \hat{X}_D \neq 0 \\
1 , \quad \sum \hat{X}_D = 0
\end{cases}
\tag{4}
\]

And the classification loss is defined as followed:

\[
L_{CLS} = - \mathbb{E} \left[ y \log \hat{y} + (1 - y) \log (1 - \hat{y}) \right] \tag{5}
\]
Defect Loss: Suppose the artificial defect map is $X_D$, and the defect map obtained by segmenting the network is $\hat{X}_D$. Use L2 distance as the loss function for defect segmentation:

$$L_{\text{defect}} = E \left[ \left( X_D - \hat{X}_D \right)^2 \right]$$

(6)

The optimization goal of the whole network can be summarized as follows:

$$L = \arg \min_{G_{\text{GAN}}, \text{defect}, \text{CLS}} \max_D \left( \omega_1 L_{\text{GAN}}(G,D) + \omega_2 L_{\text{CLS}} + \omega_3 L_{\text{defect}} + \omega_4 L_{\text{CON}}(G) \right)$$

(7)

4. Experiment

The proposed method is evaluated on two datasets including AITEX detection datasets[1] and industrial product dataset provided from the real production line. The proposed model is trained on one NVIDIA 2080 Ti GPU with images of resolution 128*128 with batch size of 32. These datasets are cropped into 128*128 sizes and the final model has a total of 73.643M parameters and implemented on PyTorch v1.4 with CUDNN v7.6.0 and CUDA v10.0.

4.1. Result

AITEX: Using the same evaluation method as the reference dataset[1], the repaired images, anomaly localizations and ground truth labels are shown in Fig 5.

$$DR = \frac{\text{Number of defective samples correctly detected}}{\text{Total number of defective samples}} \times 100\%$$

(8)

$$FAR = \frac{\text{Number of defect free samples detected as defective}}{\text{Total number of defect free samples}} \times 100\%$$

(9)

Figure 5. In figure, the first line is the pictures of different kinds of industrial products, the second line is the defect segmentation diagram generated by the segmenter branch in the proposed model, and the third line is the ground truth defect parts marked.

In this experiment, the dataset is cropped and resized to the shape as 128*128 and extracted 140 images (consisting of 45 defective samples and 95 positive samples) as the test set. The result is shown in Table 1

| Models      | DR   | FAR  | MIOU |
|-------------|------|------|------|
| Baseline[1] | 86.1%| 11.4%| -    |
As the result shown in Table 1, DCSNet got 84.4% DR, which is weaker than the result shown in reference[1]. On the one hand, DCSNet is only trained by positive sample, in the image repair stage, DCSNet generated the artificial defects by erasing pixels randomly, which is different from the real defect shape in dataset[1]. Besides, DCSNet got 4.2% FAR, which is better than reference[1], and it could have a good result in MIOU in one class learning.

4.2. Ablation experiments

We conduct an ablation study on the Industrial Product Dataset to show the effectiveness of proposed method in different backbones and structures in image repair module. Fig 6 shows the examples of the localized anomalies.

There are two evaluation indexes to measure the proposed method.

\[
overkill = \frac{False\ Positive}{all\ samples} \quad (10)
\]

\[
escape = \frac{False\ Negative}{all\ samples} \quad (11)
\]

| Backbone           | 1-overkill | 1-escape | MIOU  |
|--------------------|------------|----------|-------|
| Auto-encoder       | 0.9084     | 0.702    | 0.174 |
| Auto-encoder(GAN)  | 0.822      | 0.93     | 0.238 |
| UNet               | **0.918**  | 0.9      | 0.374 |
| UNet(GAN)          | 0.883      | **0.983**| **0.391**|

As the result shown in table 2, the backbone of image inpainting module obviously affects the performance of DCSNet. Compared with Auto-encoder structure, UNet structure can repair more details of the image in the image inpainting stage. The higher quality of the restored image, the more effective the features extracted by the segmenter and classifier, and the model could get the better score. Using GAN structure for training could make model converge faster and make the segmenter get a better MIOU, but the score of the overkill also be higher.

5. Conclusion

This paper proposed a model named DCSNet, which for the task of surface defect classification and segmentation based on One-Class learning. In the training stage, DCSNet enhances the positive sample
data by forging defects, and reconstructs the false defect through image inpainting module. DCSNet could identifies and locates the surface defect though its classifier and segmenter. In dataset [1], DCSNet got the similar performance as the supervised model, and it could achieve the detection speed of 122 FPS.

Obviously, as the result shown in ablation experiment, different backbones and the usage of GAN structures have different effects on the overkill and escape performance of the model. A new method to the trade-off between escape and overkill remains to be studied in the future.

6. References
[1]. Silvestre-Blanes, J., Albero-Albero, T., Miralles, I., Pérez-Llorens, R., & Moreno, J. (2019). A Public Fabric Database for Defect Detection Methods and Results. Autex Research Journal, 19(4), 363-374.
[2]. Liang Q, Zhu W, Sun W, Yu Z, Wang Y, Zhang D. In-line inspection solution for codes on complex backgrounds for the plastic container industry. Measurement, 2019, 148: 1-12
[3]. Yu, Z., Wu, X., Gu, X.: Fully Convolutional networks for surface defect inspection in industrial environment. In: Liu, M., Chen, H., Vincze, M. (eds.) ICVS 2017. LNCS, vol. 10528, pp. 417–426. Springer, Cham (2017).
[4]. Ding R, Dai L, Li G, Liu H. TDD-net: a tiny defect detection network for printed circuit boards. CAAI Transactions on Intelligence Technology, 2019, 4(2): 110–116
[5]. Tao Xian, Hou Wei, Xu De1. A Survey of Surface Defect Detection Methods Based on Deep Learning [J/OL]. ACTA AUTOMATICA SINICA Month:1-19[2020-10-11].
[6]. A. Abdallah, M. A. Maarof, and A. Zainal, “Fraud detection system: A survey,” Journal of Network and Computer Applications, vol. 68, pp. 90–113, Jun 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804516300571 I
[7]. B. R. Kiran, D. M. Thomas, and R. Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos,” Journal of Imaging, vol. 4, no. 2, p. 36, 2018. I, II
[8]. M. Ahmed, A. N. Mahmood, and M. R. Islam, “A survey of anomaly detection techniques in financial domain,” Future Generation Computer Systems, vol. 55, pp. 278–288, Feb 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X15000023 I
[9]. P. Bergmann, M. Fauser, D. Sattlegger and C. Steger, "MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 9584-9592, doi: 10.1109/CVPR.2019.00982.
[10]. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in neural information processing systems, 2014, pp. 2627–2680. I, III-A1, III-C1
[11]. P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.
[12]. J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu and T. S. Huang, "Generative Image Inpainting with Contextual Attention," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 5505-5514, doi: 10.1109/CVPR.2018.00577.
[13]. Yan Z., Li X., Li M., Zuo W., Shan S. (2018) Shift-Net: Image Inpainting via Deep Feature Rearrangement. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11218. Springer, Cham. https://doi.org/10.1007/978-3-030-01264-9_1
[14]. H. Liu, B. Jiang, Y. Xiao and C. Yang, "Coherent Semantic Attention for Image Inpainting," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 4169-4178, doi: 10.1109/ICCV.2019.00427.
[15]. Valvano, G., Leo, A. and Tsaftaris, S., 2020. Weakly Supervised Segmentation With Multi-Scale Adversarial Attention Gates. 13 October 2020. [Online]. Available: https://arxiv.org/abs/2007.01152.

[16]. T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10265 LNCS, pp. 146–147, 2017. I, II, III-B2, III-D, IV-C

[17]. B. R. Kiran, D. M. Thomas, and R. Parakkal, “An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos,” Journal of Imaging, vol. 4, no. 2, p. 36, 2018. I, II

[18]. “OSCT Borders X-ray Image Library, UK Home Office Centre for Applied Science and Technology (CAST),” Publication Number: 146/16, 2016. I, IV-A3

[19]. H. Zenati, M. Romain, C. Foo, B. Lecouat and V. Chandrasekhar, "Adversarially Learned Anomaly Detection," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 727-736, doi: 10.1109/ICDM.2018.00088.

[20]. J. Donahue, P. Krähenbühl, and T. Darrell, “Adversarial Feature Learning,” in International Conference on Learning Representations (ICLR), Toulon, France, apr 2017. [Online]. Available: http://arxiv.org/abs/1605.09782 II, III-A3

[21]. Akcay S., Atapour-Abarghouei A., Breckon T.P. (2019) GANomaly: Semi-supervised Anomaly Detection via Adversarial Training. In: Jawahar C., Li H., Mori G., Schindler K. (eds) Computer Vision – ACCV 2018. ACCV 2018. Lecture Notes in Computer Science, vol 11363. Springer, Cham. https://doi.org/10.1007/978-3-030-20893-6_39.

[22]. Akçay,S. A. Atapour-Abarghouei and T. P. Breckon, "Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8851808.

[23]. Xia, Y. , Zhang, Y. , Liu, F. , Shen, W. , & Yuille, A. . (2020). Synthesize then compare: detecting failures and anomalies for semantic segmentation.

[24]. M. Cordts et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 3213-3223, doi: 10.1109/CVPR.2016.350.

[25]. Hendrycks, Dan and Basart, "Steven et al. A Benchmark for Anomaly Segmentation," 2019. [Online]. Available: https://arxiv.org/abs/1911.11132.

[26]. Han Zhang, Ian Goodfellow et al. Self-Attention Generative Adversarial Networks[C]. Proceedings of the 36th International Conference on Machine Learning, PMLR 97:7354-7363, 2019.

[27]. Zhao Z., Li B., Dong R., Zhao P. (2018) A Surface Defect Detection Method Based on Positive Samples. In: Geng X., Kang BH. (eds) PRICAI 2018: Trends in Artificial Intelligence. PRICAI 2018. Lecture Notes in Computer Science, vol 11013. Springer, Cham. https://doi.org/10.1007/978-3-319-97310-4_54

[28]. P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

Acknowledgements
This work has been supported by Science and Technology Innovation Foundation of Shenzhen (JCYJ20180504165826861).