ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

A Semi-Supervised Method for Steel Surface Defect Detection Based on Soft-teacher

Heng Xiao[1,2,a,*], Changwei Zhao[1,2], Zhiyong Zhang[1,2]

[1]College of Information Engineering, Henan University of Science and Technology, Luoyang, China
[2]Henan International Joint Laboratory of Cyberspace Security Applications, Henan University of Science and Technology, Luoyang, China
[a]313445433@qq.com
[*]Corresponding author

Abstract: *Metal products are indispensable raw materials for society nowadays. Surface defects is inevitably occur during the production process, so defect detection is an important method to improve product quality. For the problem of lack of annotation of defect images in new production lines and low efficiency of manual annotation, we propose a semi-supervised defect detection method with improvement of Soft-teacher. Firstly, using mixed data enhancement to bring perturbation to the dataset and further utilize unlabeled images to enhance the effect of consistent training; then Swin-Transformer will be used as the backbone network to reduce the information loss caused by pooling, using shifted feature pyramids to do Multi-scale training, it can use the same label to supervise inputs of different sizes to obtain high-quality pseudo-labeling. Experiments on the NEU-DET dataset show that the method achieves good results in semi-supervised strip defect detection, and obtains 68.1% AP50 detection accuracy with 20% of labeled data, which is high enough to meet the needs of preliminary defect detection and labeling.*

Keywords: *Surface defect detection; Deep learning; Semi-Supervised; Object detection*

1. Introduction

Metal products are indispensable raw materials in today's world, and they are widely used in various industries, such as machinery manufacturing, automobile industry, etc[1-3]. The quality of metal products affects the product quality of its end industries. In the process of its production, surface defects will inevitably arise. These defects may affect the appearance of the products in a light way, or affect the overall performance of the products, such as corrosion resistance and service life, and reduce the rate of finished products. When defects appear in the product, the first thing to do is to reject the unqualified products, after which the production process parameters should be adjusted in a timely manner according to the type of defects. For example, during the production of hot-rolled strip steel, these six defects often appear: cracks, inclusions, patches, pitting surface, rolling-in scale, and scratches. Defect detection is a key step in controlling the quality of steel. Traditional defect detection uses manual inspection, which is time-consuming and labor-intensive, and requires a certain level of skill for workers [3-5]. For different steel products, there is a wide variety of surface defects. In addition, these defects can produce random variations due to fine tuning of the production process. The accuracy of manual inspection is not high enough, so automated high-precision inspection is needed instead of manual inspection.

With the development of technology, machine vision has been gradually applied to the field of surface defect detection. It deploys an image capture device that transmits images to an processing terminal, which in turn obtains defect category results[5]. Traditional machine vision uses image processing algorithms or manually designed classifiers for defect detection, making it hard to solve problems such as diverse defect shapes and noisy data. In recent years, object detection are commonly used for detection [5-7]. Object detection models are often use full supervised models whose accuracy relies on a large number of labeled images, but defect image labeling is very time-consuming and requires not only labeling the location but also the category. The labeling worker should have some knowledge of the type of defect to be detected, otherwise the labeling category may be incorrect.

In the field of object detection, training object detection models using a small amount of labeled data and a large amount of unlabeled data has become a research hotspot. In training object detection models, using a small amount of labeled data and a large amount of unlabeled data is called SSOD (Semi-

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

Supervised Object Detection), and the teacher-student training method is a general solution, which generally uses the teacher model to generate pseudo labeling for unlabeled images, while the student model will use labeled data and pseudo labeled images to obtain the loss, which updates the teacher weights through EMA.

To solve the problem of lack of labeling of data and low efficiency of manual labeling, this paper proposes a semi-supervised defect detection method using soft-teacher as the baseline and following the standard semi-supervised defect process, using Faster RCNN as the default detector, using mixed data enhancement to bring perturbation to unsupervised pictures and improving the effect of consistency regularization training, using Swin Transformer as the backbone network, while using a multi-scale feature training method to get rid of information loss in pooling. The detection model is validated on the NEU-DET dataset and the detection model achieves good accuracy.

2. Related Work

2.1. Surface Defect Detection

Since the late 1980s, research on defect detection based on computer vision technology has gradually developed and matured [3], where defect detection systems capture images by cameras and transmit the them to a server where a model is invoked for detection. In the early research, the model can be divided into two parts: feature extractor and classifier. Where the feature extractor extracts the features of the defect image and the classifier takes them as input to derive classification results, Zhao [4] constructed a fabric defect detection model using a multiple fractal spectrum feature extractor with SVM as classifier. Earlier defect detection methods were not robust, slow and hard to detect complex defects.

In recent years, object detection methods are commonly used for defect localization and classification. Depending on the network structure there are one-stage and two-stage detection models. The one-stage object detection network is represented by the SSD [8] and yolo [9] series as models, and the one-stage network performs both classification and localization in same operations, with the advantage of being very fast; the two-stage detection network first uses the RPN network for foreground and background classification, followed by multiclassification, with the advantage of high accuracy and good detection of dense objects, but is slower, and the representative of the two-stage network is the Faster RCNN [10].

Li [5] uses EfficientNet instead of CenterNet's backbone, proposes feature enhancement modules to enrich features, and introduces an attention mechanism to satisfy the real-time nature of the model. Haselmann [6] proposed a method that can detect surface defects with pixel accuracy, which achieves better accuracy by adding randomly generated defects to the dataset, and the method provides ideas for model training for small datasets. He[7] established a deep learning strip steel surface detection system was and the NEU-DET defect detection dataset was built to achieve high accuracy defect detection. Cheng [3] used RetinaNet for defect detection, introduced a channel attention mechanism to construct a DE-Block module in order to enable the model to detect small defects, and applied the ASFF idea to the P5 and P6 layers of FPN, and used a differential evolution algorithm to optimize the anchor frame to adapt to variable defect proportions and sizes. Hao [11] proposed a defect detection framework using MobileNet as the backbone, adding a spatial pyramid pooling module and a BiFPN module to build an improved accuracy and speed model. CABF-FCOS [12] used FCOS as the baseline and introduced an anchor free method for defect detection, in which a directional feature fusion network BFFN instead of FPN, and introduced CAM channel attention mechanism to reduce the loss of features.

2.2. Semi-Supervised research

2.2.1. Semi-Supervised Image Classification

With the rapid development of semi-supervised learning, its concepts and methods are gradually applied to the field of image classification. Semi-supervised learning methods can be divided into two categories: pseudo labeling and consistency regularization train. The pseudo labeling method uses a trained model to inference unlabeled images, and high-confidence labels as pseudo labels. However, this method has some problems, regardless of the correct class of labels, as long as the confidence level exceeds the threshold, it will be trained as pseudo label, which will lead to a large number of false samples in the pseudo label training set. Self-supervision is a specific implementation of the pseudo labeling method. Li [13] proposed a self-training approach that first trains a model trained on a small amount of labeled data, uses the model to inference pseudo labeling on unlabeled data, and later uses both labeled

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

and pseudo labeled data for training. Noisy student training [14] adds noisy data using a data augmentation method that combines labeled and unlabeled data, which expands the sample size and improves the generalization ability of the model. Consistency regularization train method means that the model makes the same judgment for small perturbations in the image. Virtual Adversarial Traning [15] proposes a virtual adversarial based loss that smooths the given label distribution and thus attenuates the effect of local perturbations on the model. Both of these methods are effective in semi-supervised learning, so FixMatch [16] combines the two methods by using weak and strong enhancement for the same image, generating pseudo labels for weak enhanced image predictions, and later using the model to inference strong enhanced images to obtain pseudo labels, using a cross-entropy loss to measure consistency. Although various semi-supervised image classification methods exist, the design concepts of these methods are not necessarily applicable in semi-supervised object detection.

2.2.2. Semi-Supervised Object Detection

Due to the high cost of manual labeling, people want to use a small amount of labeling to obtain a high accuracy model, and the research of semi-supervised object detection has been carried out gradually. Based on the research on semi-supervised image classification, the pseudo-labeling method and the consistency canonical were also used for semi-supervised object detection. STAC [17] introduced strong and weak enhancement into the training of semi-supervised models, but its teacher model could not be updated with the weight update, which limited the model performance. Mean Teacher [18] introduced the idea of EMA to update the pseudo labeling after each iteration, thus realizing the end-to-end framework. Unbiased teacher [19] uses EMA applied to object detection to update pseudo labels in real time, obtaining higher quality pseudo labels, while using Focal loss instead of cross-entropy loss to achieve robust performance in class imbalance problems. Soft-teacher [20] uses the classification scores of teachers as weights to reduce the impact of ignored objects on model performance. Pseco [21] introduced multi-scale consistency learning and noisy pseudo labeling, and experimentally evaluated the effectiveness of pseudo labeling and consistency regularization for object detection, achieving good performance and convergence speed. All of the above methods use Anchor-based detector, and DSL [22] uses an anchor free detector from an application point of view, which is easy to deploy in practice.

3. Methods

3.1. Basic Framework

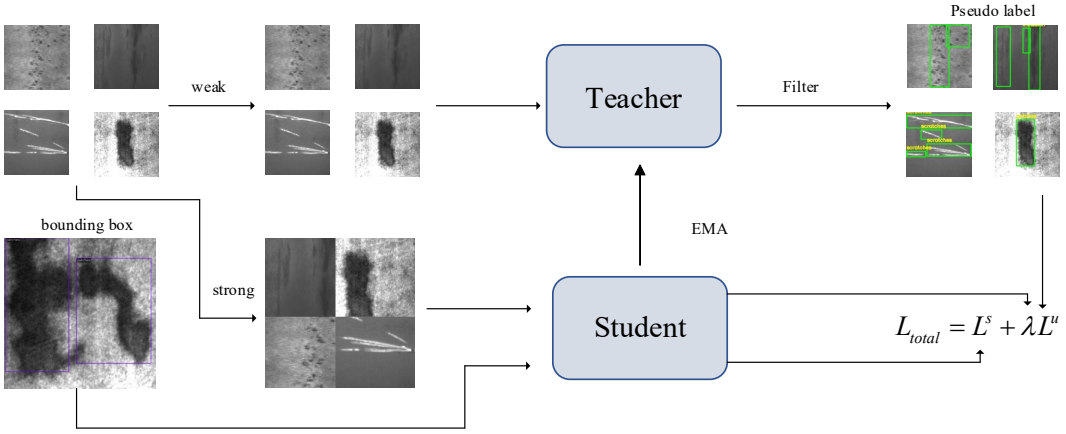

Figure 1: Basic Framework.

In this work, we make Soft-teacher [20] as a baseline, and Fig.1 shows the model training framework, which uses both pseudo labeling and consistency regularization training, using a mixed data enhance approach to add perturbations to the data, and improve the model robustness. A batch is first obtained by randomly sampling data from unlabeled data and labeled data, and the student model is supervised to get the loss, and the teacher model weights are updated by EMA (exponential moving average) in each iteration to obtain higher quality pseudo label. At the beginning of training, the weights of the teacher and student models are randomly initialized, and the teacher model is used to inference on the unlabeled data to obtain the pseudo label of categories and labels; for the student model, the labeled and unlabeled data are trained separately, with the labeled data supervised by manual labeling and the unlabeled data supervised by the pseudo-labeling of the teacher model, and thus the overall loss is obtained, and the loss

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

function is shown in Equation 1:

$$L_{total} = L^{s}_{reg} + L^{s}_{cls} + \lambda(L^{u}_{reg} + L^{u}_{cls}) \quad (1)$$

Where Ls and Lu represent supervised and unsupervised loss, and λ assigns a weight to the unsupervised loss, which is determined by the experiment. After the losses are obtained according to this formula, the weights of the teacher model are updated using the EMA method.

Since the consistency regularization affects the accuracy of the semi-supervised, generating perturbations by data enhancement can improve the generalization ability of the model. Referring to the idea of FixMatch[16], weakly enhanced and strongly enhanced perturbations are generated for unlabeled data, and the generated weakly enhanced images (horizontally flipped, resized, etc.) are sent to the teacher model, and the inference results are filtered by NMS through inference result, at which time there are still dense candidate boxes, after which the category and location pseudo labels are obtained by setting a threshold filter, after which the strongly enhanced generated images and the labeled data are are fed into the student model for inference at the same time, and the pseudo-labels obtained with labels and weak enhancement are used as supervision to obtain the final loss. Using Faster RCNN as the detector for the teacher and student models. Since the number of various defect frequencies appears unbalanced in defect detection, such as inclusions and scratches, which are significantly higher than rolled-in_scale in this used dataset NEU-DET, Focal Loss is used instead of the original cross-entropy loss.

3.2. Mixed Data Enhancement

In semi-supervised object detection, strong and weak data enhancement has been proven to be an effective training method. FixMatch uses strong and weak data enhancement for the first time, weak enhancement as image flipping and strong enhancement as color space transformation, random color erasure, etc. The weakly enhanced images are pseudo labeling by the teacher model, while the strongly enhanced images are trained by the student model with moderately complex data enhancement that can bring about data perturbation, allowing the model to learn more useful information from it. In defect datasets, grayscale images are often used and some data enhancement methods are not suitable for these dataset. We use a mixture of enhancement methods, and experimental results show that this training approach, further improves the detection accuracy of defects.

The process of Mosaic data enhancement is to determine the demarcation point of Mosaic enhancement, determine the position of each image according to the demarcation point, after that the images involved in enhancement will be Crop, if it is larger than the image position will be resize, and the label of the blended image will change with its blending method.

This mixed data enhancement application process for unlabeled data is inferred by the teacher model for unlabeled data, and the labels that are larger than the threshold value are used as pseudo-labels. Multiple data with pseudo labels are partially erased using random erasure of color blocks, after which Mosaic data enhancement is applied to them. The final pseudo label is used as the true annotation of the image.

3.3. Multi-scale training with Transformer

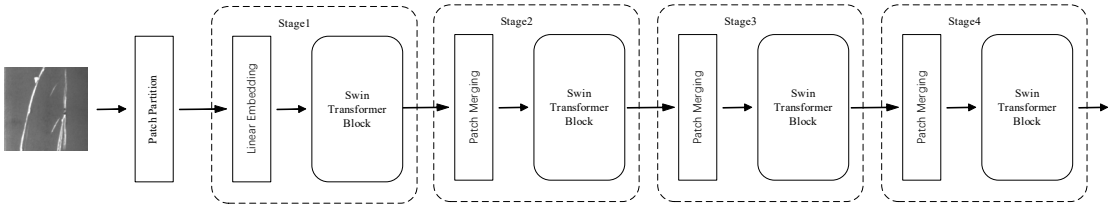

Figure 2: Architecture of Swin Transformer.

In the defect dataset, the defects vary in size and shape, and the small amount of labeled data does not represent the overall defect characteristics. This problem is solved by multi-scale training at both image and feature levels. In the detection process, multiple layers of the backbone network are used as input to the FPN. In the backbone network, pooling is usually used to downsample the images, which often results in information loss, and this problem is solved by using Transformer.

The backbone network is mainly responsible for feature extraction in defect detection, after which it is sent to the FPN for feature fusion. In the backbone network, the method of image downsampling is

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

mainly through pooling layers, which causes information loss, and in order to reduce the information loss, Swin-Transformer [23] is used as the backbone. Architecture of Swin is shown in Fig.2, which is similar to the structure of convolutional backbone network, where the image resolution is halved and the number of channels becomes two times of the original with each passing layer. Each layer includes two parts: Patch Merging and swin-transformer block. The operation of patch Merging is similar to pooling, which is using the sliding window method to take out the values at the same position in each window and put together as a new Patch, after which the concat operation is performed, which does not cause any loss of information.

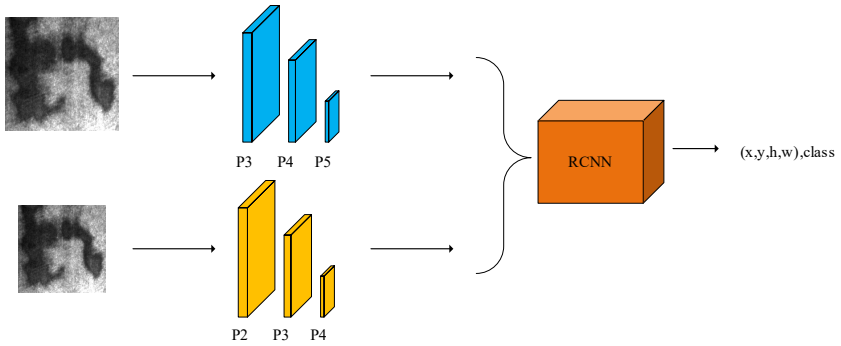

Figure 3: Multi-Scale Train.

After strong data enhancement, the images are random resized to 448 × 448 or 640 × 640 to improve the robustness of the model to defects at different scales. In the FPN， different input layers of the backbone network have different degrees of influence on the performance of defect detection accuracy. Through experiments we get the influence of each layer of the FPN in the fully supervised model, and the main ones that affect the accuracy of the model are P3, P4, P5 .To further improve the model's learning of multi-scale features, inspired by MSL [21], multi-scale consistency training at the feature level is achieved by shifting feature pyramids. The implementation steps are shown in Fig.3: the input images are 2x downsample, and the original images are a group. Since the adjacent feature maps in the FPN differ in size by a factor of 2, only the starting layer of the FPN needs to be modified, and by increasing the number of layers of the FPN, different FPN layers are used for inference for images of different sizes, and the coordinates of the same group of images do not need to be transformed after inference because the input sizes also differ by a factor of 2 That is, the same pseudo-label supervision can be used.

4. Experiments

4.1. Dataset

In this work, NEU-DET strip defect dataset is used, the images are collected from the production line, and six typical classes surface defects of hot rolled strip: crazing, inclusion, pitted_surface, scratches, patches, rolled-in_scale, as shown in Fig.4. The number of defects in each category in NEU-DET is 300, total count is 1800, the image size is 200×200, and the defects are labeled in XML format with the location and category information of each defect, there are 4200 markers in total. Firstly, the dataset is divided into training set and test set with a ratio of 7:3, after that the training set is divided into labeled data and unlabeled data, and the XML format is converted to JSON format. The dataset used in this work has 1440 images in the training set and 360 images in the test set, and 20% of the images in the training set are used as labeled data by default.

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

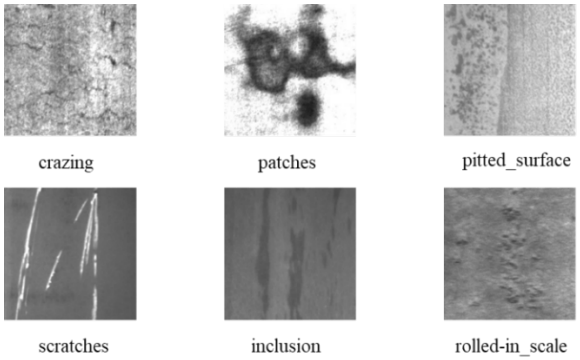

Figure 4: NEU-DET defect.

4.2. Experiments Setting

The hardware used for the experiments includes NVIDIA RTX3080 (with 12GB memory), 43GB system memory and Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz, and the software environment used is Python 3.8, Pytorch 1.10, CUDA 11.3 and cuDNN 8.2 on Linux.

The hyperparameters for the semi-supervised experiments were set as follows: 0.001 learning rate, 21000iter, 0.0001 weight decay, 5 batch size, using SGD optimizer, teacher model weight is 1, student model weight is 3, and pseudo-labeled RPN and category threshold of 0.9.

The fully supervised Faster RCNN used for the comparison was trained using 20% of the labeled data with the following hyperparameter settings: 0.001 learning rate, iteration rounds of 36 epoch, 0.0001 weight decay, backbone network loaded with ImageNet pre-training weights, batch size is 4, and SGD as the optimizer.

Using mAP as the evaluation index for this experiment, the detection frame with the cross-merge ratio IOU >0.5 and correctly predicted category was considered as a positive case, and the evaluation index of COCO was used as the evaluation method, which was mainly divided into two indexes AP50 and mAP, and the calculation formula was shown below.

4.3. Results and analysis

To achieve the optimal value of box jitter times in Soft-teacher, we employed jitter times of 2, 5, and 10 based on an unsupervised weight of 3. As demonstrated in Table 1, the experimental outcomes reveal that lower jitter times yielded the best results, while maintaining consistency with the original model's jitter scale.

Table 1: Comparison of jitter times.

Jitter times	AP50	mAP
2	66.0	29.5
5	65.7	29.1
10	65.5	29.0

To determine the appropriate value of λ for unsupervised weights in the loss function, a range of values including 1.0, 2.0, 3.0, 4.0, and 5.0 were selected for the jitter time setting of 2. Experimental results, as presented in Table 2, indicate that the optimal performance was achieved when λ was set to 3.0. It should be noted that weights greater than 3.0 resulted in slightly lower performance, which can be attributed to the random initialization of model weights at the start of the training process. Additionally, the utilization of pseudo-labeling may lead to lower accuracy in cases where the proportion of unlabeled data is significant and incorrect labeling occurs.

Table 2: Comparison of unsupervised weight.

λ	AP50	mAP
2.0	66.3	29.7
3.0	66.5	30.0
4.0	66.0	29.5
5.0	65.7	29.3

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

To demonstrate the effectiveness of the semi-supervised detection framework, we compare different labeled data percent and the performance comparison with the full supervised model Faster RCNN-FPN. As shown in Tab.3, the performance of the model improves as the percent of labeled data rises, and it can be seen that the labeled data has a great impact on the model effectiveness, so as many manual labels as possible are needed to obtain good model performance and obtain high-quality labeling. Immediately after the comparison with the fully supervised model in 20% of the labeled data, it can be seen that the semi-supervised detection framework achieves better performance compared to the fully supervised Faster RCNN.

Table 3: Comparison of percent of labeled data.

Percent of labeled	AP50	mAP
20%	66.5	30.0
40%	69.2	31.4
50%	70.6	32.0
Faster-RCNN(20%)	63.9	25.2

In order to compare the difference between mixed data enhancement and other data enhancement effects, the result shown in Tab.4. In Soft-teacher, using forms of color space conversion, geometric shape transformation and other forms of image enhancement, not applicable to grayscale images. Mixup, Cutout and Mosaic enhancement were used to compare with mixed data enhancement, in which mixed data enhancement achieved the highest accuracy, using Mixup and Cutout reduced the defect detection accuracy, and only Mosaic enhancement could further improve the detection accuracy.

Table 4: Comparison of different data enhancement methods.

Methods	AP50	mAP
base(soft-teacher)	66.5	30.0
Mixup	66.3	30.0
Cutout	66.2	29.8
Mosaic	67.3	30.3
MDE	67.6	30.5

We conducted ablation experiments to verify the effectiveness of each component. All experiments were performed with a data partition of 20% labeled data, and in Table 5, the effect of each component is shown. First is the benchmark model, which achieves 66.5% AP50 for 20% semi-supervised data. +1.1% AP50 improvement can be found when mixed data enhancement is applied to the benchmark model, due to the richer perturbation given by the mixed data augmentation approach, which enhances the generalization ability of the model. When using multi-scale training, a 0.6% improvement in AP50 can be seen, and the highest accuracy of 68.1% AP50 is obtained when both modules are used simultaneously.

Table 5: Ablation study.

MDE	MT	AP50
×	×	66.5
√	×	67.6
×	√	67.1
√	√	68.1

Comparing our proposed semi-supervised defect detection method with other SSOD methods on the NEU-DET dataset, the result as shown in Tab.6.It can be seen that the performance of STAC is not good performances when using 20% labeled data and 80% unlabeled data, which is due to the inability to update the teacher network in a timely manner, Soft-teacher and Unbiased Teacher both use the EMA method update the teacher model, so the detection accuracy is higher than STAC, while our improved soft-teacher from the perspective of data augmentation and training to achieve the highest accuracy.

Table 6: Comparison of different SSOD methods.

Methods	AP50	mAP
Soft-teacher(baseline)	66.5	30.0
STAC	60.5	25.5
Unbiased Teacher	67.5	30.2
Ours	68.1	31.0

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

5. Conclusion

In this work, we propose a semi-supervised defect detection method which is based on Soft-teacher. In practical production, labeled defect data are hard to obtain, and product quality is related to production cost. In this case, it is essential to obtain high quality labeled data to train full supervised defect detection model. In semi-supervised object detection, acquiring high quality pseudo label and consistent regularization training can significantly improve the accuracy of the model, so this paper based on these two aspects. Firstly, using mixed data enhancement method to bring perturbation in defect, it will enhance the effect of consistent training. Secondly, using Swin Transformer as the backbone network, and using multi-scale feature training to reduce information loss before the model utilizes the features to obtain high quality pseudo-labels. Experiments were conducted on the NEU-DET dataset, and the experimental results showed that our work achieved good results in semi-supervised strip defect detection, obtaining 68.1% AP50 detection accuracy using 20% of the labeled data. However, the method still has some shortcomings, as the accuracy of using Faster RCNN as a detector is also affected by the a priori anchor frame setting, and the shape and size of a small number of labeled defects in the defect data set cannot represent the whole dataset, so using an anchor free detector for semi-supervised defect detection is a research direction in the future.

Acknowledgment

The paper is partly supported by the China Nation Key R&D Program Grant No.2020YFB2008400.

References

[1] Liu, Y., Xu, K., & Xu, J. (2019). An improved MB-LBP defect recognition approach for the surface of steel plates. Applied Sciences, 9(20), 4222.

[2] Fu, G., Sun, P., Zhu, W., Yang, J., Cao, Y., Yang, M. Y., & Cao, Y. (2019). A deep-learning-based approach for fast and robust steel surface defects classification. Optics and Lasers in Engineering, 121, 397-405.

[3] Cheng, X., & Yu, J. (2020). RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection. IEEE Transactions on Instrumentation and Measurement, 70, 1-11.

[4] Zhao, C., Chen, Y., & Ma, J. (2018, October). Fabric defect detection algorithm based on MFS and SVM. In 2018 International Conference on Image and Video Processing, and Artificial Intelligence (Vol. 10836, pp. 77-82). SPIE.

[5] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

[6] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[7] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.

[8] LI F G, YILIHAMU·Yaermaimaiti. Real-time Detection Model of Insulator Defect Based on Improved CenterNet [J]. Computer Science, 2022, 49(5): 84-91.

[9] Haselmann, M., & Gruber, D. P. (2019). Pixel-wise defect detection by CNNs without manually labeled training data. Applied Artificial Intelligence, 33(6), 548-566.

[10] He, Y., Song, K., Meng, Q., & Yan, Y. (2019). An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 69(4), 1493-1504.

[11] Hao, X., Dong, T., & Zhang, D. (2021, November). A Highly Efficient Surface Defect Detection Approach for Hot Rolled Strip Steel Based on Deep Learning. In 2021 6th International Conference on Robotics and Automation Engineering (ICRAE) (pp. 318-322). IEEE.

[12] Yu, J., Cheng, X., & Li, Q. (2021). Surface defect detection of steel strips based on anchor-free network with channel attention and bidirectional feature fusion. IEEE Transactions on Instrumentation and Measurement, 71, 1-10.

[13] Li, X., Sun, Q., Liu, Y., Zhou, Q., Zheng, S., Chua, T. S., & Schiele, B. (2019). Learning to self-train for semi-supervised few-shot classification. Advances in neural information processing systems, 32.

ISSN 2616-5775 Vol. 6, Issue 3: 11-19, DOI: 10.25236/AJCIS.2023.060302

[14] Xie, Q., Luong, M. T., Hovy, E., & Le, Q. V. (2020). Self-training with noisy student improves imagenet classification. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10687-10698).

[15] Miyato, T., Maeda, S. I., Koyama, M., & Ishii, S. (2018). Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence, 41(8), 1979-1993.

[16] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., ... & Li, C. L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems, 33, 596-608.

[17] Sohn, K., Zhang, Z., Li, C. L., Zhang, H., Lee, C. Y., & Pfister, T. (2020). A simple semi-supervised learning framework for object detection. arXiv preprint arXiv:2005.04757.

[18] Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems, 30.

[19] Liu, Y. C., Ma, C. Y., He, Z., Kuo, C. W., Chen, K., Zhang, P., ... & Vajda, P. (2021). Unbiased teacher for semi-supervised object detection. arXiv preprint arXiv:2102.09480.

[20] Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., ... & Liu, Z. (2021). End-to-end semi-supervised object detection with soft teacher. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3060-3069).

[21] Li, G., Li, X., Wang, Y., Wu, Y., Liang, D., & Zhang, S. (2022, November). Pseco: Pseudo labeling and consistency training for semi-supervised object detection. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX (pp. 457-472). Cham: Springer Nature Switzerland.

[22] Chen, B., Li, P., Chen, X., Wang, B., Zhang, L., & Hua, X. S. (2022). Dense learning based semi-supervised object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4815-4824).

[23] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z. & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).