Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

LPSST: Improved Transformer Based Drainage Pipeline Defect Recognition Algorithm

Pengtao Jia, Muyuan Guo*
College of Computer Science & Technology, Xi'an University of Science and Technology, Xian, China

Abstract:

The health status of underground drainage pipelines affects the normal operation of urban drainage systems, but the current mainstream drainage pipeline defect recognition methods have the problems of poor recognition for small samples and fine-grained target defects, as well as low model generalization. Therefore, this paper proposes an improved semi-supervised transformer network based on local region proposal for drainage pipeline defect recognition (LPSST). The algorithm uses the hierarchical Swin Transformer as the backbone network, and adds a local region proposal module to the shallow network to fully pay attention to local features. Finally, a semi-supervised learning framework is used for intensive training to enhance model generalization. We studied the recognition performance of the algorithm in self-made drainage pipeline defect dataset. Compared with other mainstream algorithms such as Resnet and Efficentnet, LPSST has a recognition accuracy of 92.65% for 4 similar defects such as mismatch and deformation, exceeding 17.25% of the original backbone network. The algorithm verifies the applicability of Transformer architecture for the task of drainage pipeline defect recognition, and investigates the positive impact of semi-supervised learning on deep model training.

Keywords*: Drainage pipeline defect recognition, Transformer, Semi-supervised, Local proposal, Fine-grained image recognition.*

I. INTRODUCTION

The health status of underground drainage pipelines affects the normal operation of urban drainage systems. Accordingly, it is quite necessary to regularly test the drainage pipeline to fully understand the sound status of the pipeline [1]. In the technical method commonly used in the detection of drainage pipeline defects, professionals interpret the internal video of the pipeline captured by the inspection robot, and then classify and grade the pipeline defects. However, this defect recognition method relies too much on the inspectors' experience, which has the problems of low efficiency and high error rate, so maintenance and repair of urban underground drainage pipelines is affected. Afterwards, drainage pipeline defect recognition methods based on traditional image processing technology and deep learning model have been proposed one after another, but mostly have problems of low efficiency and high false detection rate, especially with poor recognition effect for pipeline similarity feature defects such as mismatch and deformation.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

In view of the above problems, one motivation of this paper is to use the visual Transformer architecture based on the attention mechanism as the backbone network, take advantage of its good parallelism and long-distance dependency support to perform effective modeling. A local region proposal module is introduced to fully learn global and local features of pipeline defect images. The detection algorithms based on traditional convolutional neural networks are mostly constrained by the local correlation and rotation invariance in convolution operation, which cannot better combine the local and global image features. When faced with a large amount of data, these offsets will hinder the model performance [2]. In 2020, the Google Brain team published vit [3], which for the first time applied the standard Transformer encoder directly to images, and achieved image classification effects that surpassed mainstream CNNs in ImageNet linear evaluation, thus becoming the first work of Transformer in the field of computer vision. In this work, we study the performance of the visual Transformer architecture in the task of drainage pipeline defect recognition, and the improved algorithm achieves competitive classification accuracy.

Another motivation is to avoid the label dependence of deep models, and use semi-supervised learning for intensive training to enhance the model generalization and make it more adapted to the application requirements of drainage pipeline defect recognition. The deep model performance mostly depends on the supervised learning of massive image data. However, in the practical application of the defect recognition of underground drainage pipelines, the unlabeled image data is easy to access, while labeled data relies on the manual annotation of professionals. It is a labor-intensive work to build a large-scale dataset for drainage pipeline defect detection, so it is important to learn efficiently using a small amount of labeled data. Semi-supervised learning, as a recently popular direction in the field of deep learning, requires only a small amount of labeled data to complete the relatively robust deep model training task. In this work, we study how to use unlabeled data to strengthen the pre-training model so that it conforms to the clustering hypothesis, and then a more generalizable model is established.

In order to help the engineers quickly and accurately complete the detection, analyze the defect characteristics and severity in time, this paper proposes an improved semi-supervised Transformer drainage pipeline defect recognition algorithm based on local region proposal—LPSST. The main contributions of this paper are briefly summarized as follows:

(1) The small sample category data is appropriately expanded by adjusting the image sharpness, color saturation and other color perturbation data enhancement methods, which alleviates the problem of unbalanced categories in the self-made drainage pipeline dataset.

(2) A local region proposal module is introduced into the shallow network of the backbone network to improve local feature extraction. By setting a reasonable mask, the self-attention range is limited to a specific local area, so that equivalence calculation is possible when the number of windows is not significantly increased.

(3) By using a semi-supervised learning framework that integrates consistency regularization and

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

pseudo-labels for intensive training of the model, the model accuracy and generalization are greatly improved, demonstrating remarkable experimental effect.

II. RELATED WORK

2.1 Based on Traditional Vision Algorithm

With the continuous development of computer vision technology, scholars at home and abroad attempt to use traditional computer vision algorithms to determine drainage pipeline defects. In 2002, Fieguth et al. segmented pipeline images based on morphological methods to extract geometric features, and then recognized images using fuzzy neural networks [4]. In 2008, Ming-Der Yang team from Taiwan performed wavelet transform processing on pipeline images to gather co-occurrence matrix and texture features. Then, machine learning methods were used for automatic detection of drainage pipeline defects [5]. In 2012, Motamedi et al. performed grayscale, filtering, and morphological preprocessing operations on drainage pipeline images, which provided a theoretical basis for non-destructive testing of urban drainage pipeline defects [6]. Kirstein S et al. integrated the shortest path algorithm, Hough straight line transform algorithm and Canny edge detection method to detect drainage pipeline defects [7]. In 2014, based on the prior information and visual features related to the inner surface cracks of the drainage pipeline, Mahmoud et al. adopted Sobel edge detection method to detect the candidate crack edges, thus fulfilling the detection of drainage pipeline crack defects [8]. In 2016, in view of characteristics of industrial-grade drainage pipeline defects, Mayuri et al. converted the RGB image of the pipeline into a grayscale image and then completed the detection and recognition based on the extracted industrial pipeline diameter defects and structural defects [9]. In 2017, Zhonghu Li et al. proposed an image edge detection algorithm based on back-propagation neural network (BP), which was used to detect the edge of the corrosion defect image for the inner wall of the drainage pipeline [10].

Drainage pipeline defect detection technology based on traditional computer vision has achieved certain research results, but there are still problems such as high requirements for input image quality, too complicated preprocessing, low generalization, low recognition rate, and single detection defect category.

2.2 Based on Deep Learning

With the development of deep learning-based image technology theory, researchers have gradually applied deep learning models to the field of drainage pipeline defect detection. In 2018, Jack C.P. Cheng et al. proposed a drainage pipeline defect detection method based on Faster R-CNN [11] (Region-Convolutional Neural Network) [12], which allows accurate detection of pipeline cracks, sediments, infiltration and other defects. Moreover, it was pointed out that, by increasing dataset size and number of convolutional layers, it is possible to effectively improve the algorithm performance. In 2019, Dirk Meijer et al. from the Netherlands proposed a classification network structure based on Convolutional Neural Network (CNN) and applied it to defect detection of drainage pipelines [13]. In the same year, Bing Lv et al. proposed a CNN-based intelligent detection method for drainage pipeline defects in closed-circuit

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

television (CCTV) video frames [14]. In 2020, Maohui Zheng et al. used Genetic Algorithm (GA) to optimize the Extreme Learning Machine (ELM) neural network, which provided a new data-driven modeling method for the recognition and diagnosis of urban drainage pipeline defects [15]. In 2021, Qianqian Zhou et al. proposed an intelligent detection and classification method for drainage pipeline defects based on convolutional neural networks [16].

The drainage pipeline defect detection and recognition method based on deep learning model improves the recognition rate for pipeline defect types. However, due to the influence of the deep model itself, there are still problems such as slow detection rate and low accuracy in fine-grained defect detection.

Regarding the problem of low accuracy in feature detection under limited discrimination and low generalization of the model with limited label data amount, this paper studies how to integrate global and local effective feature information, and uses a semi-supervised reinforcement training framework for effective learning of a small amount of label data to establish a high generalization model and achieve accurate recognition of drainage pipeline defects with similar characteristics.

III. MAIN METHODS

3.1 Model Structure

The overall structure of the LPSST algorithm is shown in Fig 1. With Swin Transformer [17] as the backbone network, the algorithm is improved based on the structural defect image characteristics of the drainage pipeline. The algorithm pre-training model uses the color perturbation-based data enhancement method to expand the small sample category label data, integrates the Transformer encoder based on feature area screening to perform feature extraction. The detection head recognizes and outputs the defect category. After the unlabeled data is weakly/strongly perturbed, it is input to the pre-trained model, and the intensive training is completed by virtue of two semi-supervised learning methods of pseudo label and consistency regularization. Finally, the original pre-trained model is updated with the trained parameters to establish the final recognition model.

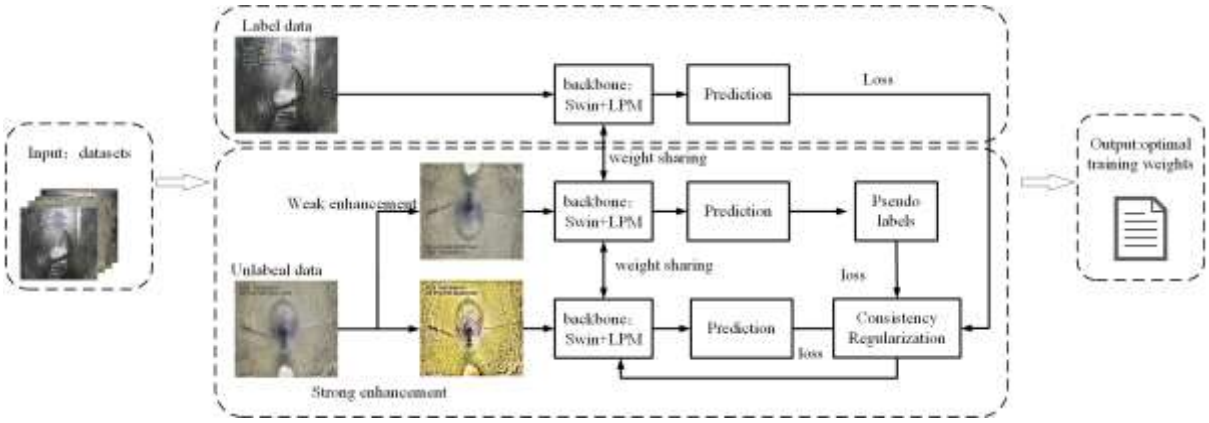


Fig 1: A schematic diagram of the model structure based on LPSST algorithm

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

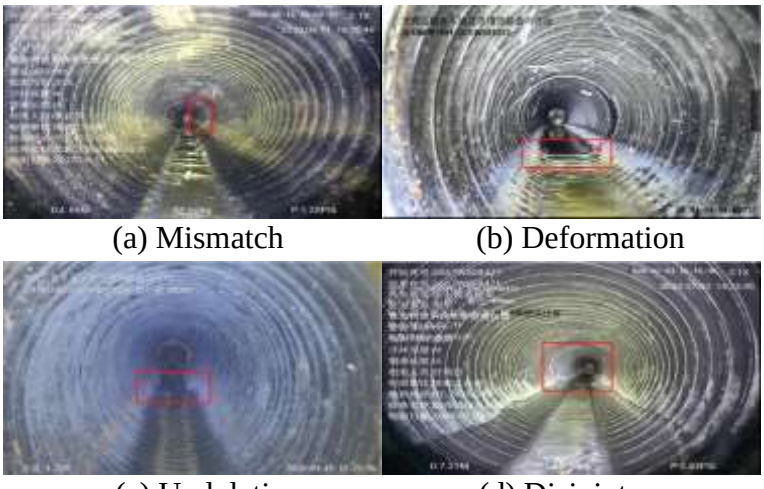
3.2 Selection of Backbone Network

As one of the mainstream backbone networks in the current Vision Transformer field, Swin Transformer displays performance comparable to mainstream CNNs [18] in various image tasks. The network introduces the hierarchical idea of CNNs and proposes a cascaded Transformer, which merges image blocks layer by layer in depth to build a pyramid structure. Particularly, the prominent contribution of the network is the design of shifted window attention mechanism. By setting a reasonable mask, the self-attention range is limited to non-overlapping local windows, which allows equivalent calculation without significant increase in the number of windows. The layered architecture of Swin Transformer demonstrates multi-scale modeling flexibility. Suitable for image classification and dense detection tasks, it has linear computational complexity for image size, which can evaluate the performance in downstream tasks such as object detection and semantic segmentation.

In this work, Swin-T, a lightweight version of Swin Transformer, is used as the backbone network. Swin-T has similar complexity to ResNet-50 [19], which guarantees a balance between model speed and accuracy. which keeps the balance between the model computation speed and accuracy.

3.3 Local Region Proposal Module

In drainage pipeline detection, it is found that the structural defects of some drainage pipelines have highly similar characteristics. Moreover, amid pipeline image acquisition, the camera position angle is relatively fixed, making it impossible to capture a full-angle image around a defect. Therefore, images collected by drainage pipeline defect detection often have different defect types and highly similar features. As shown in Fig 2, the four images represent four kinds of drainage pipeline structural defects: mismatch, deformation, undulation, and disjointness. Defect characteristics displayed by such images are slightly different, making misrecognition easily occur in recognition of deep learning model network. Easily affected by the experience of the inspector, manual recognition is prone to wrong classification.



(a) Mismatch (b) Deformation

(c) Undulation (d) Disjointness

Fig 2: Drainage pipeline defect categories with similar characteristics

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

According to the image feature representation of drainage pipeline structural defects, highly similar feature defects mean that the model network must capture more subtle feature differences to achieve accurate defect recognition. This paper draws on the idea of designating local areas for fine-grained feature extraction in fine-grained visual classification, which is to combine global features to form the final feature representation, enhance local and global correlation, and improve classification accuracy. The local area proposal module is designed and added in Swin-Transformer, and the specific structure diagram is shown in Fig 3.

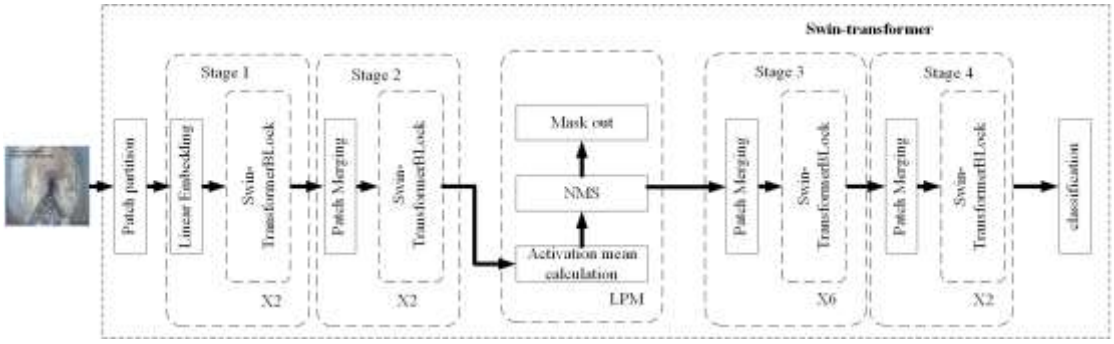


Fig 3: Backbone network structure integrating local proposal module

The activation average $\bar{\alpha}_{\omega}$ is calculated based on the window feature map A_{ω} mapped to each window. As shown in Equation 1, H_{ω}, W_{ω} represent the window height and width, respectively.

$$\bar{\alpha}_{\omega} = \frac{\sum_{x=0}^{W_{\omega}-1}\sum_{y=0}^{H_{\omega}-1} A_{\omega}(x,y)}{H_{\omega} \times W_{\omega}} \tag{1}$$

After calculating the activation average $\bar{\alpha}_{\omega}$ of all windows, all the $\bar{\alpha}_{\omega}$ is sorted. A higher $\bar{\alpha}_{\omega}$ value corresponding to the window means that the window area contains more information. The background of drainage pipeline defect image is mostly pure color inner wall of drainage pipeline. The information is relatively monotonic, so window selection is performed based on non-maximum suppression (NMS). The input of NMS is all possible predicted bounding box (Equation 2) and a given iou threshold, while the output is the predicted bounding box filtered by the NMS algorithm (Equation 3).

$$predictions = \left[\left[x\_max, x\_min, y\_max, y\_min, score\right], [*], \cdots, [*]\right] \tag{2}$$

$$result = \left[x\_max, x\_min, y\_max, y\_min, score\right] \tag{3}$$

NMS filters out windows with higher $\bar{\alpha}_{\omega}$ values for retention, constructs key areas, and uses

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

reasonable mask output to mask the discarded window areas in screening. It then establishes a new feature image of the same dimension for input to the next stage of the network. The specific module process is shown in Fig 4.

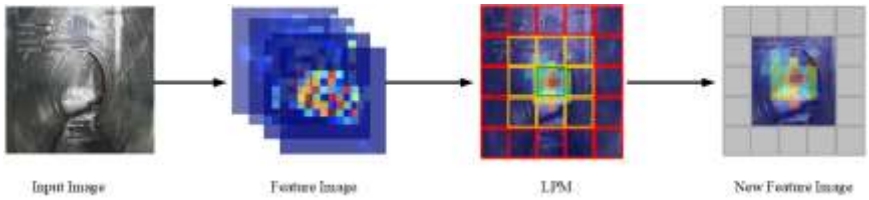


Fig 4: Local region proposal module process

3.4 Semi-supervised Reinforcement Training

Fix Match [20] is a semi-supervised learning method proposed by the Google Brain team. Different from other semi-supervised methods, Fix Match uses cross-entropy to compare the unlabeled data with weak enhancement and strong enhancement for consistency regularization.

The core of consistency regularization is to hope that a sample and its disturbed samples will get similar outputs after the classifier processing, mainly by maintaining consistency in predictions before and after unlabeled data enhancement (disturbance), so that the learned decision boundary is located in a low-density area, and no specific label is required. By constructing an unsupervised regularization loss term between the perturbed prediction result Y and the normal prediction result y on the unlabeled data, the model generalization ability is strengthened. Its expression is shown in Equation 4.

$$D\left[p_{model}\left(y \mid Augment(x), \theta\right), p_{model}\left(Y \mid Augment(x), \theta\right)\right] \tag{4}$$

Where, D is the metric function, generally KL divergence or JS divergence is used, and cross entropy is used in this chapter. Augment (·) is a data augmentation function used to add some noise perturbations, with θ representing model parameters.

The semi-supervised training of Fix Match performs normal supervised learning on labeled data. The loss function L_S adopts the common standard cross-entropy loss function, and its mathematical expression is shown in Equation 5:

$$L_S = \frac{1}{B}\sum_{b=1}^{B} H\left(p_b, p_m\left(y \mid \alpha(x_b)\right)\right) \tag{5}$$

Where, B is the batch size of the labeled samples, x_b is the training sample, and p_b is the one-hot label.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

$p_m(y \mid x)$ is used as category distribution predicted by the model network input x. $\alpha(\square)$ Represents the form of weak enhancement used.

For the unlabeled data, the reinforcement training of the model is completed through the following four steps:

(1) Weak and strong enhancement of unlabeled data;

(2) Use the trained model to make predictions on the augmented samples. For weakly enhanced samples, if the highest prediction probability $Q_b = \arg\max(q_b)$ of the predicted result exceeds the set threshold, it is considered as a valid sample, and the result is assigned to the sample as a pseudo-label. Otherwise, the sample is ignored.

(3) For strongly enhanced samples, the prediction results output by the model and the pseudo-labels are used for loss calculation to obtain the loss function L_u. The specific calculation process is shown in Equation 6.

$$L_u = \frac{1}{\mu B}\sum_{b=1}^{\mu B}\left(max\left(q_b\right) \geq \boldsymbol{\tau}\right) H\left(Q_b, p_m\left(y \mid \mathbf{A}\left(u_b\right)\right)\right) \tag{6}$$

Where, $\boldsymbol{\tau}$ represents the scalar hyperparameter of the threshold. When $max(q_b) \geq \boldsymbol{\tau}$, retain the pseudo-label of the corresponding data.

(4) Calculate the final loss function $Loss = L_s + \lambda L_u$ of the model. λ Is a fixed scalar hyper parameter. Thus, reverse gradient propagation is performed on $Loss$ to complete the update of the entire model network.

IV. EXPERIMENTS

4.1 Dataset

Pipeline closed-circuit television detection (CCTV) technology is so far one of the most widely used technologies in the detection of urban underground drainage pipelines. It has been widely used in status detection of rainwater pipelines and sewage pipelines, which runs through the whole stage of pipeline construction and acceptance [21]. This paper conducts experiments on the self-made dataset for CCTV drainage pipeline defect recognition.

In order to make full use of the training samples to prevent over-fitting, regarding the problem of long-tailed dataset presented by self-made drainage pipeline dataset, data enhancement is performed to

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

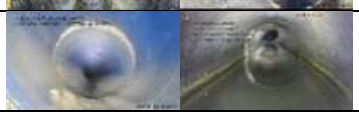
expand the data of small sample categories, achieve the balance between classes and reduce the offset probability of the model network during training. Considering that most of the drainage pipe defect images are irrelevant backgrounds, and that defects only occur in small local positions, operations such as simple random rotation, flipping cannot achieve a good expansion effect.

Therefore, by adjusting the four values of image sharpness, brightness, color saturation, and contrast and combining them randomly, this paper uses color perturbation-based data enhancement to increase or decrease some color components in the color space. The experimentally validated method can achieve efficient data enhancement without loss of details of drainage pipeline defects. The effect after data enhancement is shown in Fig 5.

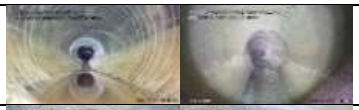
Fig 5: comparison of renderings after data enhancement (the upper left shows the original image, the upper right shows the contrast enhancement, the lower left shows the color saturation enhancement, and the lower right shows the brightness reduction)

The four types of drainage pipeline structural defects, including mismatch, disjointness, deformation, and undulation, are sorted and the small sample data categories are expanded until a unified number to form a drainage pipeline defect recognition dataset of 2,500 images per category, with an image size of 800 × 800. Table I shows the relevant information of the four types of structural defect images.

TABLE I. Drainage Pipeline Defect Recognition Dataset

CATEGORY NO.	DEFECT CATEGORY	CATEGORY ABBREVIATION	QUANTITY	IMAGE EXAMPLE
0	Deformation	BX	2500	
1	Mismatch	CK	2500	

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

2	Undulation	QF	2500	
3	Disjointness	TJ	2500	

4.2 Comparison of Experimental Results and Analysis

In order to verify the reasonable validity of the LPSST algorithm, several variants of the improvement process are investigated in the comparative experiments in this section, including improvement based on local region proposal and semi-supervised learning. The purpose of this setting is not to compare methods directly, but to evaluate whether our improvement can effectively enhance model performance.

Under the same experimental conditions, six groups of LPSST models were comparatively tested respectively, namely Swin-Transformer, Resnet50, Densenet [22], Efficientnet [23], and the backbone network Swin-Transformer-LPM with local region proposal module, and finally integrated with semi-supervised reinforcement training. The comparison is based on four parameters: the highest accuracy rate (TopAcc@1), the average accuracy rate (Mean Acc), the number of parameters (Param), and the computational complexity (FLOP). The specific experimental results are shown in TABLE II.

TABLE II. Comparison of recognition results of different networks

MODEL NETWORK	**TOPACC@1**	**MEANACC**	**PARAMS（M）**	**FLOPs（G）**
CNN-based models				
Resnet50	70.00%	67.90%	25.26	3.53
Densenet	71.70%	69.20%	7.98	2.79
Efficentnet	71.10%	68.60%	5.28	0.39
Transformer-based models				
Crossformer	73.40%	70.80%	30.7	4.9
Focal-Transformer	74.30%	70.90%	29.1	4.8
Swin-Transformer	75.40%	71.01%	28.28	4.35
Swin-Transformer-LPM	**76.20%**	**71.94%**	**29.21**	**4.49**
Semi-supervised training				
LPSST	**92.65%**	**91.86%**	**27.52**	**4.34**

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

Resnet-SSL	91.25%	89.26%	25.26	3.53

According to the data analysis in TABLE II, compared with the traditional convolutional network, the improved original backbone network Swin-Transformer achieves the best accuracy results. Compared with ResNet with the same complexity structure, small growth is exhibited in both the parameter amount and computational complexity, possibly because the small data amount is insufficient to display advantages of massive data in the Transformer architecture. After local region proposal module is added, the model accuracy is further improved. After intensive training of the Swin-Transformer-LPM pre-training model with the help of the semi-supervised learning framework, the final LPSST model achieves the best accuracy rate of 92.65%, with a linear improvement of 17.25% compared with the original Swin-Transformer. At the same time, the parameter number and computational complexity are slightly reduced. In addition, ResNet model incorporating semi-supervised learning experiences a significant increase in performance. In summary, the comparative experimental results fully demonstrate the reasonable effectiveness of the LPSST algorithm improvement. The algorithm achieves a substantial improvement in the classification and recognition performance, and verifies the high representation ability of semi-supervised learning.

4.3 Ablation Experiment

In order to investigate the influence of certain variables (or methods) on the model performance, this section still selects the same dataset for ablation research. One parameter is changed for each ablation, and the final parameter values of the pre-training model are maintained constant in the remaining parameter configurations.

(1) The proportion of the number of samples with and without labels

In order to better investigate the influence of the number of labeled samples on the performance of the semi-supervised learning model, relevant ablation research was carried out, with the number of labeled samples set as: 1000, 3000, 6000, 9000. The number of unlabeled samples was kept at 9000, and comparison test was carried out to investigate the influence of different proportions of samples with and without labels on the recognition results. The specific results are shown in TABLE III.

TABLE III. Experimental Comparison of Different Labeled Sample Sizes

NUMBER OF LABELED DATA	**PROPORTION OF LABELED SAMPLES**	**TOPACC@1**	**MEANACC**
1000	1：9	79.32%	75.80%

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

3000	1：4.5	84.82%	80.40%
6000	**1：2.5**	**92.65%**	**91.86%**
9000	1：0.1	82.83%	79.67%

According to the data analysis in Table 3, when the proportion of labeled samples is 1:2.5 (6000 labeled samples and 9000 unlabeled samples), the LPSST model has the best recognition effect, with an accuracy rate of 92.65%. For its reason, a small proportion of labeled data will affect the generation of pseudo-labels, which will affect model training and reduce model performance. A great proportion of labeled data will reduce the generalization performance of the model, leading to overfitting in model training.

(2) Optimizer

Model instability appears during the training process, possibly because the training of the semi-supervised learning method is quite sensitive to the optimizer selection. In order to reduce the instability and seek the best training strategy for the LPSST algorithm, in this section, different optimizer types are selected as hyperparameters, and the experimental configurations of other models are kept unchanged to investigate the impact of training in different ways on model performance.

By default, we use Adam W [24] as the optimizer, which is a common choice for training vit models [3, 17, 25]. In addition, two common SGD and LARS optimizers are used for experiments. The specific experimental results are shown in TABLE IV.

TABLE IV. Experimental comparison of different optimizers (the number of labeled data is 6000)

OPTIMIZER	**TOPACC@1**	**MEANACC**
SGD	79.32%	78.34%
LARS	82.88%	81.65%
AdamW	**92.65%**	**91.86%**

According to analysis of the results in Table 4, under this training condition, the model trained by the optimizer Adam W achieves the best recognition accuracy, and the SGD commonly used in the CNN model performs poorly in the Transformer backbone network. Hence, Adam W is finally selected as the training optimizer for the LPSST algorithm in this paper.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

V. CONCLUSION

This paper designs an improvement scheme from three perspectives: small sample category data expansion, enhancement of local and global feature information capture, and semi-supervised learning. Finally, the drainage pipeline defect recognition algorithm was established with the improved semi-supervised Transformer based on local proposal (LPSST). Experiments show that LPSST can achieve high-precision recognition of four fine-grained structural defects of drainage pipelines, with recognition performance significantly improved.

The deep learning method is applied to the intelligent detection and recognition of urban underground drainage pipeline defects, and satisfactory detection and recognition results have been achieved, but there are still problems such as slow detection rate and model performance susceptible to the label data quality. In the later period, we plan to further complete the algorithm optimization from the aspects of deep model lightweight and active model learning ability. It is hoped that this simple baseline will inspire researchers to reconsider the role of semi-supervised representation learning in practical engineering applications.

REFERENCES

[1] Hu Y (2019). Research on intelligent detection technology of underground drainage pipeline defects based on deep learning. Xi 'an: School of Civil Engineering and Architecture, Xi'an University of Technology.

[2] D 'Ascoli, S., Touvron, H., Leavitt, M., Morcos, A., Biroli, G., & Sagun, L. (2021). Convit: improving vision transformers with soft convolutional inductive biases. International Conference on Machine Learning. PMLR, 2021: 2286-2296.

[3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., & Houlsby, N. (2020). An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[4] Fieguth, P. W., & Sinha, S. K. (2002). Automated analysis and detection of cracks in underground scanned pipes. Proceedings of 1999 International Conference on Image Processing (Cat. 99CH36348). Kobe, August 06, 2002. IEEE, 1999, 4:395-399

[5] Yang, M. D., & Su, T. C. (2009). Segmenting ideal morphologies of sewer pipe defects on CCTV images for automated diagnosis. Expert Systems with Applications, 36 (2p2), 3562-3573.

[6] Motamedi, M., F Faramarzi, & Duran, O. (2012). New concept for corrosion inspection of urban pipeline networks by digital image processing. Conference of the IEEE Industrial Electronics Society. IEEE.

[7] Kirstein, S., Muller, K, Walecki-Mingers, M, & Deserno, T. M. (2012). Robust adaptive flow line detection in sewer pipes. Automation in Construction, 21 (Jan.), 24-31.

[8] Halfawy, M. R, & Hengmeechai, J. (2014). Efficient Algorithm for Crack Detection in Sewer Images from Closed-Circuit Television Inspections. Journal of Infrastructure Systems, 20 (2): 04013014.

[9] MD Shinde, & Wane, K. (2016). An Application of Image Processing to Detect the Defects of Industrial Pipes. International Journal of Advanced Research in Computer and Communication Engineering, 2016, 5 (3): 979-981.

[10] Li Z, Zhang L, Yan J. (2017) Research on image edge detection algorithm for pipeline corrosion visual measurement. Journal of Electronic Measurement and Instrumentation, 2017, 31(11):1788-1795.

[11] Ren, S., He, K, Girshick, R, & Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39 (6), 1137-1149.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
September-October 2022 Page No. 151-164
Article History: Received: 05 April 2022, Revised: 27 April 2022, Accepted: 03 May 2022, Publication: 14 May 2022

[12] Cheng, J. C. P, & Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. Automation in Construction, 95 (NOV.), 155-171.
[13] Meijer, D, Scholten, L, Clemens, F, & Knobbe, A. (2019). A defect classification methodology for sewer image sets with convolutional neural networks. Automation in Construction, 104(AUG.), 281-298.
[14] Lv B, Liu Y, Ye S, Yan Z. (2019). Convolutional-neural-network-based sewer defect detection in videos captured by CCTV. Bulletin of Surveying and Mapping, 2019 (11): 103-108.
[15] Zheng M, Liu S. (2021). Defect diagnosis of urban drainage pipelines based on GA optimized ELM neural network. Journal of Harbin Institute of Technology, 2021, 53 (05):59-64.
[16] Zhou Q, Situ Z, Teng S, Chen G. (2021). Intelligent Detection and Classification of Drainage Pipe Defects Based on Convolutional Neural Networks.China Water & Wastewater, 2021, 37(21):114-118.
[17] Liu, Ze, et al. (2021) "Swin transformer: Hierarchical vision transformer using shifted windows."arXiv preprint arXiv: 2103. 14030.
[18] X Chen, Fan, H, Girshick, R, & He, K. (2020). Improved baselines with momentum contrastive learning. Preprint arXiv: 2003. 04297, 2020.
[19] K. He, X. Zhang, S. Ren and J. Sun, (2016). "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp.
[20] Sohn, K., Berthelot, D., Li, C. L., Zhang, Z., Carlini, N., & Cubuk, E.D., et al. (2020). Fixmatch: simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 2020, 33: 596-608.
[21] Chen X. (2019). Application of CCTV detection technology for municipal pipelines. Water Conservancy and Hydropower Construction, 2019, 38(1): 102-105.
[22] Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., & Keutzer, K. (2014). Densenet: implementing efficient convnet descriptor pyramids. Eprint Arxiv.
[23] Tan, M, & Le, Q. V. (2019). Efficientnet: rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR, 2019: 6105-6114.
[24] Wang, J, Wu, Z, Chen, J, & Jiang, Y. G. (2021). M2tr: multi-modal multi-scale transformers for deepfake detection. arXiv preprint arXiv: 2104. 09770 (2021).
[25] Touvron, H, Cord, M, Douze, M, Massa, F, & H Jégou. (2020). Training data-efficient image transformers & distillation through attention. International Conference on Machine Learning. PMLR, 2021.
[26] Zhang, F., Li, M., Zhai, G., Liu, Y. (2021). Multi-branch and Multi-scale Attention Learning for Fine-Grained Visual Categorization. In, et al. MultiMedia Modeling. MMM 2021. Lecture Notes in Computer Science (), vol 12572. Springer, Cham. https://doi.org/10.1007/978-3-030-67832-6_12.
[27] Mayuri, D. S. and Prof. Kishor W., (2016). "An Application of Image Processing to Detect the Defects of Industrial Pipes", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016, pp 979-981.
[28] Wang, C. Y., Liao, H., Wu, Y. H., Chen, P. Y., & Yeh, I. H.. (2020). CSPNet: A New Backbone that can Enhance Learning Capability of CNN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE.