Detection of Rail Surface Defects Based on CNN Image Recognition and Classification

Lidan SHANG, Qiushi YANG, Jianing WANG, Shubin LI, Weimin LEI

School of Computer Science & Engineering

Northeastern University(NEU), Shenyang, China

shang_lidan@yeah.net, qsyeung@gmail.com, anjingdemuguan_g@163.com, lsb_neu@126.com

Corresponding Author: **leiweimin@ise.neu.edu.cn**, Tel: +86 13940235647

***Abstract*—Due to the rapid advances in railway industry, the rail surface defect detection task which inspects whether the rail is defective has become an increasingly critical issue. Detecting rails by an automatic and swift approach instead of present manual inspections enables the work more efficient and safe currently. In this paper, we propose a novel two-stage pipeline method for rail defect detection by localizing and classifying rail images. Specifically, in the first stage, we get the cropped images which focus on the rail part instead of the whole-original image by integrating traditional image processing methods. In the second stage, we put the cropped images into a fine-tuned convolution neural network (CNN) and extract part-level features for rail images classification. Especially, in the rail image defect detection scenario, we should take recall into account to some extent, so we propose a novel loss function to leverage both of them in the second stage. The results show that the proposed method has strong robustness and achieves practical performance in defect detection precision.**

***Keywords*— Rail defect detection, Rail image classification, Object localization, Part-based CNN, Recall**

I. Introduction

With the high-speed railway's rapid development, the security issue brings people's attention. Although the continuous improvement of rail technology decreases the occurrence of the internal defects, the frequent rail fatigue wear cracks [15], scars and other surface defects caused by poor environment may accelerate the deterioration of the rail. That may induce the formation of rail internal defects, even contribute to rail fracture and train derailment accident. Therefore, the timely rail surface detection is the best prevention from rail accidents.

So far, the traditional defect detection methods adopted in this field can be summarized into the following three categories: 1) manual inspection which is usually subjective and does not produce an auditable visual record 2) physical detection methods like eddy current detection method [1][16] whose signal is hard to deal with due to various external interferences and ultrasonic detection [2][18] which cannot guarantee the real-time of signal processing and may leave detection blind spot. 3) computer vision methods [5][6] for object detection like CNN [3][4][40][41] which takes up large computing resources to localize object. Moreover, these methods don't take the recall rate which is regarded as a crucial evaluation parameter for object detection into account.

In this paper, we propose a two-stage pipeline algorithm based on object localization and convolution neural network (CNN) for object detection. The object localization which is independent of CNN aims to localize the rail image through rail image segmentation. While the CNN is introduced to extract rail surface features from the cropped rail images and then divides the rail images into two categories—intact and defective. Additionally, a novel loss function is proposed to leverage the recall and precision of classification results because the specific rail surface defect detection task requires high recall to ensure more defective rail images are found. Based on the algorithm, the recall of rail surface defect detection is improved under the premise of high precision. Moreover, the training time and the calculation complexity are reduced simultaneously.

The rest of this paper is organized as follows. Section II reviews the previous approach of the defect detection. Detailed implementation of our method is shown in Section III. The training environment and comparative results are presented in section IV. The remaining paper is dedicated to the experiments on our rail image dataset and conclusion.

II. Related Work

In defect detection, ultrasonic [14][15], eddy current [16], infrared [17] and laser [18][19] methods are widely used. Ultrasonic testing is an important branch in surface detection which uses the distinctive reflection characteristics on different media interfaces to detect defects. However, in practical applications, due to the immature technology, the requirements for the detected objects are extremely high but the accuracy is low. Eddy current detection relies on the change of voltage and

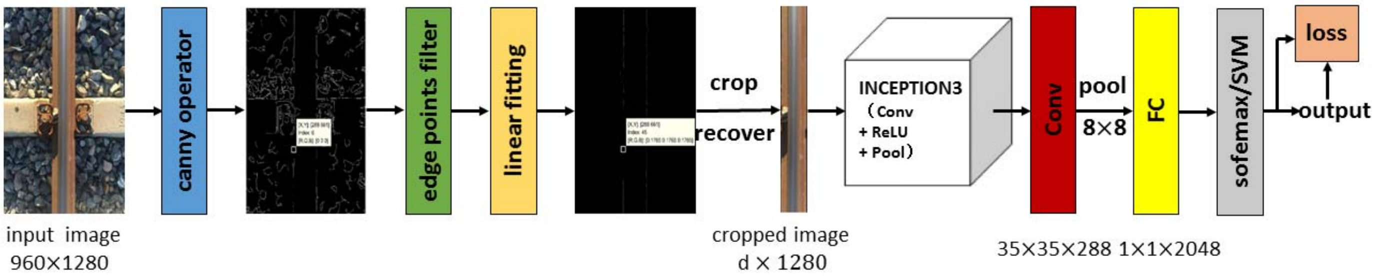


Figure 1. Architecture of our model

impedance of detection equipment which is caused by strength and distribution of eddy current. But its effect is influenced by object surface state and detection standard. Infrared detection method bases on changes in radiation that is related to the temperature distribution of detected objects. While its effect is easily affected by the ambient temperature. All approaches above are labor-intensive and they extremely rely on many artificial standards. Our methods detect rail defect with images and train a convolution neural network to learn high-level feature from images and to classify rail automatically without any manual standard.

CNN as a special deep neural network model has been demonstrated [20] to give superior performance in the ImageNet challenge [21], we have witnessed a resurgence of research interests in CNN [22]-[25]. Therefore, our work also turns to CNN for automatic feature extraction and classification CNN's particularity is manifested in two aspects: (1) the connection between its neurons is not fully connected. (2) The weight of the connection between some neurons in some layer is shared . LeNet-5 [30] is the most basic network structure of CNN, which is a good solution to the problem of handwriting recognition. Alexnet [29] structure comes up with the ideas of Relu, dropout and data augmentation [29], which solves the problem of nonlinear and over-fitting. Since then, the development of deep learning is greatly promoted. Network structures such as VGG, NiN, GoogleNet [26]-[28] have been proposed one after another and considerable results have been achieved on the corresponding data sets. Inception-v3 [12] is a branch of GoogleNet.

Data imbalance [10][31] is a commonplace issue in the field of rail defect detection. In the actual rail defect detection, since intact rail images are more readily available than defect images, there is a difference between the two categories. In the training of CNN, the majority class samples tend to be paid more attention to, while the minority class samples may be relatively neglected. Attributed to this operating mechanism, it will cause the distortion of classification boundary, the decline of precision and the wrong conclusion. In order to solve those problems, there are generally two ways: the up-sampling and down-sampling. Up-sampling [31][32] such as SMOTE [37] and ADASYN [33] synthesizes new minority class samples, based on the similarity of feature spaces. But when the sample data of the original dataset is too big, the synthesized data may degrade performance. Down-sampling such as EasyEnsemble [35][36], One-Sided Selection [34] balances positive samples with negative samples in their feature space by deleting the number of majority class samples, which may cause the loss of feature information. Instead of synthesizing or deleting samples, we add a parameter consisting of recall and precision to loss function so that the algorithm can give consideration to both precision and recall. To a certain extent, it can overcome the problem of performance degradation and information loss.

III. PROPOSED METHOD

A. *Overview*

In this section, we propose a two-stage approach to rail inspection based on the combination of traditional object localization algorithm and CNN. Our method is designed to extract high-level rail surface features and classify the rail as defective or intact. The overview pipeline is shown in Figure 1.

B. *Object Localization*

In order to realize object localization, we adopt the pre-treatment of rail images, the removal of false edge points, and the line fitting method. Due to the full use of the geometric characteristics of the rail itself, our object localization algorithm is simple and efficient.

***1) Rail Image Preprocessing*:** Given the rail geometric characteristics, we propose a robust and less computational algorithm for rail localization. We first use the weighted average method to transform the original colour images into grayscale images and use the adaptive median filtering [5] which can adjust the size of filter window based on the noise points to reduce noise. Next, we use the Canny edge detector to obtain the edge points and save the column of the edge points. Then we further remove false edge points among the saved edge points by using the ratio d/l of the rail in the images (where d and l indicate the width of rail and the width of the grayscale image as shown in Figure 3). Details are described as follows.

2) Rough Removal of False Edge Points: We first calculate the column difference between every two adjacent edge points in a row obtained by Canny edge detector from left to right. The difference is denoted as $\mathrm{x_i} = \mathrm{c_{i+1}} - \mathrm{c_i}$, where c_{i+1} and c_i represent the column of right and left edge point of the two respectively. If x_i satisfies the roughly dynamic range ($0.8\frac{\mathrm{d}}{\mathrm{l}} < x_i < 1.2\frac{\mathrm{d}}{\mathrm{l}}$), we save c_i in left edge point group and

(a) (b) (c)

Figure 2. **(a)** is the edge detection result used Canny edge detector; **(b)** is the edge detection result with our linear fitting method; **(c)** is the result of rail localization

c_{i+1} in right point group respectively. This operation should be performed row by row and it end up with two group edge points eventually. The method is designed to exclude some large deviant false edge points obtained after Canny edge detector.

3) Further Removal of False Edge Points: Calculate the standard deviation $'\mathrm{s}'$ of column saved in the above two groups. If $'\mathrm{s}'$is relatively large, that indicates there still remain partial false edge points. Then we propose the constraint condition$\left(1-\frac{\mathrm{c}}{\mathrm{s}}\right)\mathrm{n} < x_i < \left(1+\frac{\mathrm{c}}{\mathrm{s}}\right)\mathrm{n}$ to adaptively narrow the dynamic range and further remove the false edge points from our preserved edge points. ($'\mathrm{n}'$ represents the columns of the input image). The parameter c in the formula can also be adjusted to achieve optimal results. The effect of c on the results of the classification will be presented in the experiment part.

4) Linear Fitting: After "Further removal of false edge points", the real edge points may be removed, thus, we need to restore them. We use the existing two sets of edge points for linear fitting respectively to get the intercept and slope. Next, the real edge points can be restored by the intercept and slope approximately. So we can achieve the object localization by real edge points and recovered edge points.

The edge detection results using Canny edge detector and our linear fitting method are illustrated in Figure2-(a) and the Figure2-(b) respectively. Based on the mentioned method, the rail localization has been achieved accurately, as shown in Figure2-(c).

C. Classification via CNN

In this section, we leverage a fine-tuning CNN to pre-train images and propose a new loss function to balance precision with recall.

1) Transfer Learning Based on Inception-v3: After the object localization, we use transfer learning and CNN to achieve the two-category images. Due to local receptive field and parameter sharing methods, CNN can achieve the effect of automatic feature extraction. In order to complete the initial feature extraction, we use the Inception-v3 which has an impressive performance in CNN [8] as our basal network. There are four reasons for choosing Inception-v3: (1) Inception-v3 allows the rail images which have been processed to be spatially converged in low dimensions without worrying about losing a lot of feature information. (2) The feature map of the rail slowly decreases during the propagation, and the number of filters increases before the pooling to reduce the loss of information and the probability of bottlenecks. (3) The large convolution is decomposed into small convolution. In this way, the feature parameters of the rail defects are further reduced, and the convolution kernel parameters and complexity are reduced to complete the transition from a compact network to a sparse network. (4) Transfer learning which is based on the mature model such as inception-v3 can make up for the shortcoming of lack of rail samples and reduce the computational complexity.

Transfer learning [7][9] is a learning method which uses existing knowledge to solve problems in different but related fields and relaxes two basic assumptions in traditional machine learning in order to migrate existing knowledge to deal with learning problems in the target area where there is only a small quantity of tagged sample dataset. There are many kinds of classification methods of transfer learning, but the method of transfer learning between images can be mainly divided into the following two categories [11]. (1) Map the corresponding math relationship to expand the amount of data in the target domain when the difference between the target domain and the source domain is small. (2) When there is a large difference between the target domain and the source domain, transfer learning is built on the convolution neural network layer and accomplishes the migration of features. There is a distinct difference between the features of the rail dataset and the features of the inception

Figure 3. Diagrammatic sketch of d and l

dataset, so what we do belongs to the second category.

***2) Balance* of *Recall* and *Accurac*y:** On considering the issue about class imbalance as well as the special characteristics of rail defects detection task, it is necessary to leverage the recall and precision of classification results. A common method for addressing the choice of classification-task evaluation index is to introduce a variable F-score, which is the weighted harmonic mean between recall and precision. Besides, many methods [37] for classification via CNNs use cross entropy as the loss function without any part responding to recall directly. Thus, on the particularity of our task, we propose to add the F-score to the cross entropy loss.

The precision, recall and weighting factor of the last m samples are denoted by P, R and β respectively. We focus on some latest samples' precision , recall and F-score whose numbers are all the square of m because only the latest samples during training period play a most part in the results of present stage. Moreover, we define intermediate iteration index, current iteration index, estimated probability for corresponding class and the F-score parameter as i , m , p , $F_{i-score}$ respectively.

$$\begin{cases} R^{(m)} = \dfrac{1}{\lfloor\sqrt{m}\rfloor} \displaystyle\sum_{m-\lfloor\sqrt{m}\rfloor}^{i\le m} R^{(i)} \\ P^{(m)} = \dfrac{1}{\lfloor\sqrt{m}\rfloor} \displaystyle\sum_{m-\lfloor\sqrt{m}\rfloor}^{i\le m} P^{(i)} \end{cases} \tag{1}$$

The cross entropy loss $CE^{(i)}$ and the F-score parameter $F^{(i)}{}_{i-score}$ are as follows:

$$F_{i-score}^{(i)} = \frac{(1+\beta)P^{(i)}R^{(i)}}{\beta^2 P^{(i)} + R^{(i)}} \tag{2}$$

$$CE^{(i)} = -\frac{1}{m}\sum_{i=1}^{i\le m}\left[p^{(i)}\log\left(p^{(i)}\right) + \left(1-p^{(i)}\right)log\left(1-p^{(i)}\right)\right] \tag{3}$$

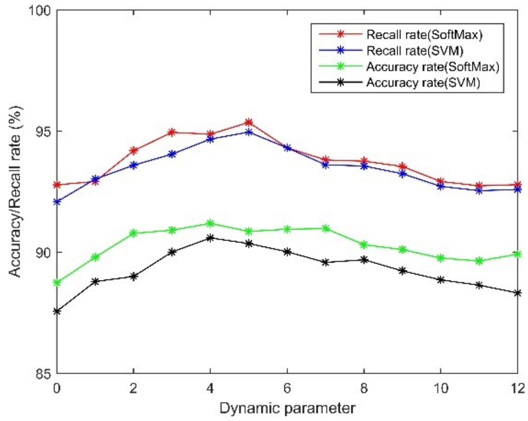


Figure 4. The influence of dynamic parameter c on accuracy and recall

$$F_{i-score}^{(i)} = \frac{(1+\beta)P^{(i)}R^{(i)}}{\beta^2 P^{(i)} + R^{(i)}} \tag{4}$$

In order to normalize those two parts, we use Z-score standardization algorithm [17] to get their normalization forms $\overline{CE}^{(i)}$, $\bar{F}_{i-score}$ and write them as:

$$\begin{cases} \overline{CE}^{(i)} = \dfrac{CE^{(i)} - \mu}{\sigma} \\ \bar{F}_{i-score} = \dfrac{F_{i-score} - \mu_F}{\sigma_F} \end{cases} \tag{5}$$

In the above, we define μ , σ as the mean and variance of sequence $CE^{(i)}$ and μ_F , σ_F as the mean and variance of $F_{i-score}$. The sequence of $CE^{(i)}$ consists of all trained samples' cross entropy and the sequence of $F_{i-score}$ consists of the last training samples' F-scores whose number is the square of m. The proposed loss function with above parts is defined as:

$$\text{Loss} = \overline{CE} - \bar{F}_{i-score} \tag{6}$$

It is intuitive that the F-score with both precision and recall is a strong constraint for loss function so that the constraint enables the classifier to learn towards the correct direction.

IV. Experiments

We evaluate our method on our dataset that consists of 2 classes with defective rail and intact rail images. The models are trained on the 5793 training images with 5327 intact rail images and 466 defective rail images and evaluated on the 2276 validation images with 2033 intact rail images and 243 defective rail images. Then we get the final results on the 1517 test images. Every rail image in the dataset is each frame from the rail videos shot by the high-speed cameras located in front of special locomotives. Moreover, our experiments are conduct with Intel i5-5200 processor 2.20GHz and 8G RAM.

For the first stage, our baseline is the results produced by original images trained in the Inception-v3 with proposed loss function where β is 2. For the second stage, our baseline is the

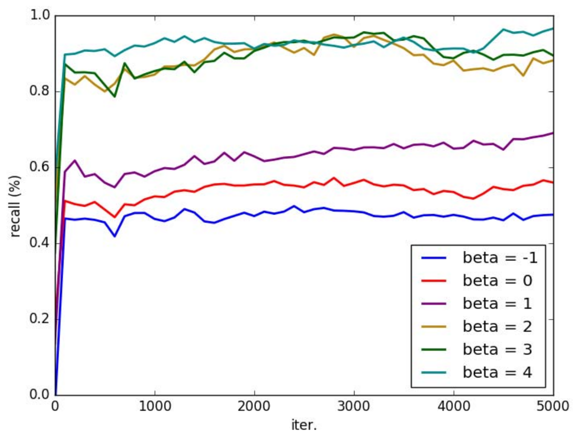


Figure 5. The results of recall with different β on the training images set

results produced by original images trained in Inception-v3 with cross entropy loss function. We evaluate the recall rates and F-scores of our methods and baseline.

A. Implementation Details of Inception-v3 Net

In this part, our work is mainly divided into two steps, the first step is to take full advantage of inception- v3 for pre-training. G.Hiton and Y.LeCun *et al.* [38][39] show that low-level image features are typically captured at lower convolution layers. Higher convolution layers capture more advanced features. The final fully-connected layer is thought to capture information related to solving the corresponding task. The second step is to modify the full connection layer, combine the transfer learning with the network, and migrate trained weights to the corresponding part of the modified network.

The Inception-v3 model has a total of 46 layers consisting of 11 inception modules. There are 96 convolution layers which are mostly combined in parallel. We keep the parameters of the original convolution layer trained in Inception-v3 model, only replace the last fully connected layer and train the new mono-layer neural network. The reason why the layer is reserved is that the fully-connected layer can act as a firewall in transfer learning [12]. In particular, when the source and target domains differ greatly, the fully-connected layer can maintain the capability of the model. Then we add a dropout layer behind the connected layer to prevent over-fitting and set the value as 0.5 [11].

The input of the network is 960 × 1280 preprocessed images. After being trained by Inception-v3 model, the images input the 35 × 35 × 288 convolution layer and then enter the 8 × 8 × 288 pool layer for further feature compression. Finally, the images input 1 × 1 × 2048 fully -connected layer for training new parameters. In our experiments, the network learning rate is 0.01 and the batch-size and iteration number are set to 64 and 5000 respectively.

B. Performance of Object Localization Modules.

Using the new loss function ($\beta = 2$) which will be illustrated in Subsec.D, we perform a comparative experiment on the performance of object localization. The result of the experiment is shown in Figure 4. It turns out that both accuracy and recall roughly increase with c first and then decrease. From the experiment results, when c is 5, the recall and accuracy reach 95.36% and 90.86% respectively under softmax, while 94.96% and 90.34% under SVM. It reflects that the softmax classifier outperforms SVM, in our experiment, the recall and accuracy cannot reach the maximum at the same time, but considering the particularity of our task, we choose c as 5 for later classification task.

TABLE 1. PERFORMANCE OF THE PROPOSED OBJECT LOCALIZATION METHOD (UNDER SOFTMAX CLASSIFIER)

Method	Recall (%)	Accuracy (%)
Original images-CNN	92.54	88.27
Cropped images-CNN ("c=5")	95.36	91.19

To further verify the performance of our localization algorithm, we feed original images and cropped images into convolution neural network followed by softmax separately. The result is shown in Table 1. It indicates that, when c is 5, the accuracy and recall of image classification increase by 2.82% and 2.92% compared with the raw image.

It's reasonable to assume that the improved performance comes from the better learned features without extra inference in rail background clutter. And this further corroborates that our localization strategy is crucial for CNN to learn more representative features for classification.

C. Training Rail Images Classification.

The normalization parameter $\bar{F}_{i-score}$ changes are related to index β, so we cannot regard $\bar{F}_{i-score}$ as the compositive index among different β. We evaluate the changes of recall and precision respectively with different values of β. Besides, the baseline is $\beta = -1$. As β increases, the recall increases but the precision decreases. The changes of recall are shown in Figure 5 and precision are shown in Figure 6. We can observe from figures that when β is over 2, the recall increases slowly but the precision decreases quickly. When β is 2, as shown in Figure 6, it is a critical value for precision because of a sudden drop.

D. Comparison with the Results of Test Set.

After being trained as above shown, the classifier achieves good performance on the test set. We sweep over different choices of the parameter β, the test results of recall and precision are shown in Table 2. As the similar issue we argue in the training stage, the precision decreases quickly when β is over 2 and recall increases slowly. Thus we note that β=2 is the best choice for both recall and precision in loss function.

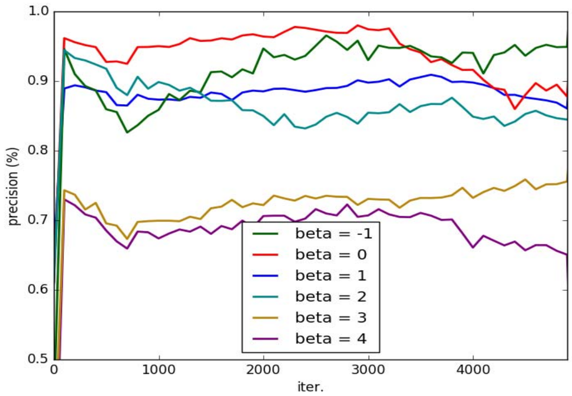


Figure 6. The results of precision with different β on the training images set

TABLE 2. RESULT OF TEST SET

β	Recall	Precision
-1(baseline)	47.69	96.21
0	50.23	95.33
1	64.28	94.64
2	**92.54**	**92.08**
3	92.77	72.57
4	93.16	65.86

V. CONCLUSION

In this paper, our two-stage pipeline method based on object localization and CNN is tested on our rail image dataset and it is validated that both the recall and precision of our rail image classification receive an impressive and robust performance compared with raw rail image owing to the introduction of the new loss function and object localization algorithm. The results also indicate that our localization strategy is significantly for CNN to learn more representative features for classification. Besides, the utilization of transfer learning strategy also reduces the computational complexity. It is expected that the method can be used for actual rail defect detection automatically.

However, our object localization method is specifically designed for railway image given the rail geometric characteristics and this detection method cannot guarantee the real-time work. In future work, we will also devote to further performance optimization on the classification precision and recall and try to achieve real-time multi-category goal which can identify the specific defection type.

REFERENCES

[1] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567, 2015.

[2] Cui Xuehong , Liu Yun ,Wang Chuanxu , Li Hui . Defect Classification for tire X-ray images using convolutional neural network [M]. Electronic Measurement Technology, 2017,40(5):168-173

[3] Pang Na1, Zhao Qilin1, Rui Ting1, Zhu Bin2 . Current Situation and Development of Bridge Detection Technology Based on Machine Vision [D]. Modern Transportation Technology, 2015,12 (6):25-31

[4] X. Gibert, V. M. Patel, D. Labate, and R. Chellappa, “Discrete shearlet transform on GPU with applications in anomaly detection and denoising,” EURASIP Journal on Advances in Signal Processing, vol. 2014, no. 64, pp. 1–14, May 2014

[5] DENG Xiu-qin , XIONG Yong , PENG Hong. Effective adaptive weighted median filter algorithm. Computer Engineering and Applications，2009， 45（35）: 185-187.

[6] J.J. Cunningham, A.E. Shaw, and M. Trosino, “Automated track inspection vehicle and method,” May 2000, US Patent 6,064,428.

[7] Fei-Fei L, Fergus R, Perona P. One-Shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4).

[8] Zhuang FZ, Luo P, He Q, Shi ZZ. Survey on transfer learning research. Ruan Jian Xue Bao/Journal of Software, 2015,26(1):26-39 (in Chinese). http://www.jos.org.cn/1000-9825/4631.htm

[9] Ze Ping.Yang Research of unbalanced data based on neural networkshttp://www.jos.org.cn/1000-9825/4631.htm

[10] Chu Jinghui, Wu Zehuan, Lu Wei, Li Zhe.Diagnosis System of Breast Tumors Based on Migration Learning and Deep Convolution Neural Networks [J] .Journal of Laser and Optoelectronics, 2018,55 (05): 051004.

[11] Chen-Lin Zhang, Jian-Hao Luo, Xiu-Shen Wei, Jianxin Wu In Defense of Fully Connected Layers in Visual Representation Transfer

[12] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1):321-357.

[13] Wu H, Hayes M J, Weiss A, et al. An evaluation of the Standardized Precipitation Index, the China-Z Index and the statistical Z-Score[J]. International Journal of Climatology, 2001, 21(6):745-758.

[14] Dr. rer. nat. Dr.-Ing. E. h. Josef Krautkrämer. Ultrasonic Testing of Materials[J]. 1984, 51(1):218-224.

[15] Marais J J, Mistry K C. Rail integrity management by means of ultrasonic testing[J]. Fatigue & Fracture of Engineering Materials & Structures, 2010, 26(10):931-938.

[16] Yang R, He Y, Gao B, et al. Lateral heat conduction based eddy current thermography for detection of parallel cracks and rail tread oblique cracks[J]. Measurement, 2015, 66(1/2):54-61.

[17] Greene R J, Yates J R, Patterson E A. Crack detection in rail using infrared methods[J]. Optical Engineering, 2007, 46(5):051013.

[18] NakHyeon Kim, Hoon Sohn, SoonWoo Han. Rail Inspection Using Noncontact Laser Ultrasonics[J]. Detection of Committee, 2012, 32(32):696-702.

[19] Zhao Y, Sun J H, Ma J, et al. Application of the hybrid laser ultrasonic method in rail inspection[J]. Insight - Non-Destructive Testing and Condition Monitoring, 2014, 56(7):360-366(7).

[20] Ting Sun, Lin Sun, Dit-Yan Yeung. Fine-Grained Categorization via CNN-Based Automatic Extraction and Integration of Object-Level and Part-Level Features ☆[J]. Image & Vision Computing, 2017, 64.

[21] A. Krizhevsky, I. Sutskever, G. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C. Burges, L. Bottou ,K. Weinberger(Eds.), Advances in Neural Information ProcessingSystems25,CurranAsso-635 ciates, Inc., 2012, pp. 1097–1105.

[22] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: visualising image classification models and saliency maps, arXiv preprint arXiv: 1312.6034.

[23] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, DeCAF: a deep convolutional activation feature for generic visual recognition,565 in: Proceedings of the International Conference on Machine Learning,2014,pp.647–655.

[24] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks,in: Proceedings of the European Conference on Computer Vision,2014,pp.818–833

[25] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild ,in: Proceedings of the IEEE International Conference on ComputerVision,2015,pp.3730–3738

[26] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C}. IEEE Conference on and Pattern Recognition(CVPR),2015:Computer Vision 1-9.

[27] LeCunY ,Bottou L, BengioY ,Haffner P. Gradient-based learning applied to document recognition [C]// Proc. IEEE, 1998. USA: IEEE,199R:227R-2324

[28] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, Andrew D Back. Face Recognition: A Convolutional Neural Nehvork Approach [J]. IEEE Trans. on Neural Nettvorks (S1045-9227), 1997, 8(1): 98-113.

[29] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 2012.

[30] LeCun Y. LeNet-5, convolutional neural networks[J]. URL: http://yann. lecun. com/exdb/lenet, 2015.

[31] H. He, E. A. Garcia, Learning from unbalanced data, IEEE Transactions on Knowledge and Data Engineering, 2009, 21(9): 1263-1284.

[32] N. V Chawla,over-same ling K.W. Bowyer, L.O. Hall, et al., SMOTE: Synthetic minority technique, Journal of Artificial Intelligence Research, 2002 16: 321-357.

[33] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for unbalanced learning, IEEE International Joint Conference on Neural Networks (IJCNN),2008: 1322-1328.

[34] M. Kubat, S. Matwin, Addressing the curse of unbalanced training sets: One-sided selection, The 14th International Conference on Machine Learning, 1997: 179-186.

[35] J. Zhang, I. Mani, KNN approach to unbalanced data distributions: A case study involving information extraction, ICML'2003 Workshop on Learning from Imbalanced Data Sets, 2003.

[36] X. Y Liu, J. Wu, Z. H. Zhou, Exploratory under Sampling for class imbalance learning, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2009, 39(2):539-550.

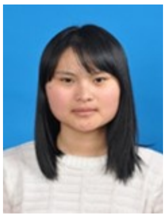
[37] Kline D M, Berardi V L. Revisiting squared-error and cross-entropy functions for training neural network classifiers[J]. Neural Computing & Applications, 2005, 14(4):310-318.

[38] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.

[39] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard,W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural computation, 1989

[40] Uijlings, Jasper RR, et al. "Selective search for object recognition." International journal of computer vision 104.2 (2013): 154-171.

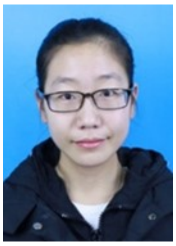
[41] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,"2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 580-587.doi: 10.1109/CVPR.2014.81.



Lidan Shang was born in Henan Province, China in 1996. She is an undergraduate student at Dept. of Communication Engineering in Northeastern University in China. Her current research interests are in the area related to Computer Vision, Image Processing

Qiushi Yang was born in Jilin, China in 1996. He now majors in communication engineer in Northeastern University. His current interests are deep learning and image processing.

Jianing Wang was born in Hebei,China in 1995.He now majors in communication engineering in Northeastern University .He is interested in the field of deep learning ,communication theory and internet technology.



Shubin Li was born in Jiaozuo, China in 1997.She majors in Electronic and Information Engineering in Northeastern University. Her current researches are in area related to Digital Image Processing and Defect detection.

Weimin Lei received the B.E. and M.E. degree in computer software from Nankai University and Chinese Academy of Sciences in 1992 and 1995 respectively, and the Ph.D. degrees from Dalian University of Technology in 1999. He is currently working as a professor in Institute of Communication and Information System at Northeastern University of China. He has published more than 80 papers. His recent research interests include real-time multipath transmission optimization, real-time signal processing, and future network architecture.