OPTOELECTRONICS LETTERS Vol.19 No.5, 15 May 2023

Damage identification technology of substation instrument panel for solving imbalanced classification*

YAO Nan[1], WU Xi[2], ZHAO Yuxi[1], SHAN Guangrui[1], and QIN Jianhua[1]**

1. Research Institute, State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210000, China

2. State Grid Wuxi Power Supply Company, Wuxi 214000, China

(Received 9 October 2022; Revised 8 January 2023)

Edge computing plays an active role in empowering the power industry as a key technology for establishing data-driven Internet of things (IoT) applications. Traditional defect diagnosis mainly relies on regular inspection of equipment by operation and maintenance personnel at all levels, and its accuracy relies on the human experience. In actual production, the image data of some dashboard damage types are easy to collect in large quantities, while some dashboard damage types occur less frequently and are more difficult to collect. The use of edge computing nodes allows flexible and fast collection of smart meter data and transmission of the reduced data or results to a cloud computing center. In this study, we provide a fresh balanced training approach to address the issue of learning from unbalanced data. In the equilibrium training phase, a new impact balance loss is introduced to reduce the influence of samples on the overfitting decision boundary. Experimental results show that the proposed balance loss function effectively improves the performance of various types of imbalance learning methods.

Document code: A **Article ID:** 1673-1905(2023)05-0296-5

DOI https://doi.org/10.1007/s11801-023-2169-z

In the power system, there are inevitably many power equipment need to be exposed to the external environment for a long time, bearing the task of high load and high voltage work. It will also be affected by bad weather, which is very easy to make the equipment appear different degrees of damage, and inevitably will make its various performance weakened. With the deepening of the intelligence of the power system, the number of terminal device connections has increased dramatically, and the traditional cloud computing has been unable to carry the massive data processing. The effective combination of edge computing[1,2] and power system not only provides strong support for the safe operation of power system, but also brings new opportunities and challenges for the edge intelligence of power system. While the massive amount of data generated at the edge of the power grid brings convenience to people, there are also problems of security risks. On the one hand, it is how to use deep learning algorithms to detect abnormal equipment images and obtain valid information from them to enhance the power system. On the other hand, the current processing of massive data relies entirely on cloud computing platforms, which can lead to the problem of idle resources on end devices. Edge computing is a key link in establishing data-driven power system applications with advantages such as real-time efficiency, network stress relief, and intelligent security. And when there is a problem of damage to the power equipment, the edge computer system can be used for reasonable distribution, without affecting the operational efficiency of the whole system.

Resampling, cost-sensitive learning[3], and adjustable learning rates[4] are a few strategies that have been suggested in recent years to lessen the effect of data class imbalance on model performance. These techniques haven't been proven to work with deep learning models, hence they only work with shallow models. Deep representational learning[5] can likewise be confused by the same variables. In this study, we provide a fresh balanced training approach to address the issue of learning from unbalanced data. We also propose an ideal knowledge transfer wide residual network fragmentation image classification method to address the issue of low resolution acquired images and poor visual information. It results in many participants in the teacher network and low accuracy in the student network, as well as to compensate for the shortcomings of using size models for classification tasks. The teacher network and student network are modeled after edge computing and edge intelligence, respectively[6,7]. The teacher network with the best knowledge transfer directs the training of the student network using a combination of a hidden layer attention transfer algorithm

* This work has been supported by the Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd.: Research on Intelligent Diagnosis Technology of Substation Tour View Spectrum Based on Edge Computing (No.J2021066).

** E-mail: yaonanyn123@163.com

and an output layer knowledge distillation algorithm. The student network may manage a variety of computational tasks independently without sharing the underlying data with the instructor network thanks to such a distributed edge computing node. Additionally, edge computing's decentralized and hierarchical computing architecture offers the network more dependable deep learning calculations. Our suggested approach can be easily combined with other recent resampling, cost-sensitive learning, and adaptive learning rate methods to address class imbalance, because our losses are not restricted to particular tasks, models, or training methods, and edge computing also has richer data and application scenarios.

The trained deep learning-based transformer breakage detection model detection is the current widely used inspection method[8]. The weighted samples that cause deep neural network (DNN) overfitting for extremely unbalanced data training are reduced using a novel loss-sensitive method that we describe in this study. We create a brand-new impact-balancing loss and weight samples differently based on how they affect the boundary information. In particular, using the reciprocal of the effect of each sample, we recalculate the weight loss proportion. Impact balancing fine-tuning and conventional training are the two components of our approach. The proposed impact balance loss reduces the impacts that cause the boundary information to be overfit during the fine-tuning stage.

Without removing the data and retraining the model, the influence function enables us to predict the change in model parameters when eliminating samples. Let $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})$ denote the model parameterized by $\boldsymbol{w}$ with n training data $(x_1, y_1)\ldots(x_n, y_n)$, where x_i is the i-th training sample with label y_i.

We rescale the weights in the opposite proportion to the sample impact during the fine-tuning step to correct the imbalance. If the distribution of training data at a position $(\boldsymbol{x}, \boldsymbol{y})$ is slightly altered, the impact of that point can be approximated by parameter changes. This is how the influence function is displayed as

$$I(\boldsymbol{x},\boldsymbol{w}) = -H^{-1}\nabla_{\boldsymbol{w}}\boldsymbol{L}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w})). \tag{1}$$

The derivation from $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{w})$ affects the equilibrium loss function, and since the inverse Hessian matrix must be thoroughly calculated for the vector $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{w})$, direct use is practically not possible. Influencing the equilibrium loss function from the middle derivative is almost impossible to use it directly since the inverse matrix must be extensively computed for a vector. As a result, we change $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{w})$ to a straightforward and efficient influence balancing weighting factor. First, $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{w})$ can be completely disregarded because we only need to consider the relative impact of the training samples rather than their absolute value. This is due to the fact that all training samples are typically multiplied by the inverse of Hessian. The impact equilibrium loss weighting factor is then created as

$$IB(\boldsymbol{x};\boldsymbol{w}) = \left\|\nabla_{\boldsymbol{w}}\boldsymbol{L}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}))\right\|_1, \tag{2}$$

where $IB(\boldsymbol{x}; \boldsymbol{w})$ is the magnitude of the gradient vector. It can be further simplified when using softmax cross-entropy loss. We consider the overfitting of model boundaries to be very important, so we focus on the changes in the last fully connected (FC) layer where the deep neural network works. Set $\boldsymbol{h}=[h_1,\ldots,h_L]^{\mathrm{T}}$ as the hidden feature vector as the input of FC layer and $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{w})=[f_1,\ldots,f_K]^{\mathrm{T}}$ as the output of $f_K := \sigma(\boldsymbol{w}_K^{\mathrm{T}}\boldsymbol{h})$, where δ is the softmax function. The weight matrix of the FC layer is denoted by $\boldsymbol{w}=[w_1,\ldots,w_K]^{\mathrm{T}}\in R^{K\times f}$. Ultimately, the gradient of the loss is calculated by the following equation:

$$\frac{\partial}{\partial w_{kl}}\boldsymbol{L}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w})) = (f_k - y_k)h_l. \tag{3}$$

The cross-entropy loss[9] with function or the mean square error (MSE) loss applied for regression both yielded the same results. The inverse term of which can be used to reweight the factors in order to reduce the weights of the influential samples during the fine-tuning process:

$$\begin{aligned} IB(\boldsymbol{x};\boldsymbol{w}) &= \sum_k^K\sum_l^L\left|(f_k - y_k)h_1\right| = \\ &\sum_k^K\left|(f_k - y_k)\right|\sum_l^L\left|h_1\right| = \\ &\left\|\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}) - \boldsymbol{y}\right\|_1 \cdot \left\|\boldsymbol{h}\right\|_1 . \end{aligned} \tag{4}$$

Finally, the equilibrium loss is given by the following equation:

$$\boldsymbol{L}_{IB}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w})) = \frac{\boldsymbol{L}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}))}{\left\|\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}) - \boldsymbol{y}\right\|_1 \cdot \left\|\boldsymbol{h}\right\|_1}. \tag{5}$$

Additionally, we modify the impact equilibrium loss in Eq.(5) by adding a class reweighting term λ_k

$$\boldsymbol{L}_{IB}(\boldsymbol{w}) = \frac{1}{m}\sum_{(x,y)\in D_m}\lambda_k\frac{\boldsymbol{L}(\boldsymbol{y},\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}))}{\left\|\boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}) - \boldsymbol{y}\right\|_1 \cdot \left\|\boldsymbol{h}\right\|_1}. \tag{6}$$

Since normalization is done and n_k is the number of samples in class k in the training dataset, λ_k has a similar scale for each class. α is introduced as an adjustment hyperparameter. The following two results result from ranking the weights. First, by slowing down majority loss minimization, k lessens the bias of the decision boundary caused by the overall distribution imbalance. Second, λ_k controls the wisely adjusted weights of the samples according to the classes to which the samples with high influence belong. Because the influence of the minority sample is very substantial due to the dearth of data, if the sample falls into the minority category, λ_k is less than the majority sample and has minimal negative weighing on the loss.

The student's network has insufficient precision in recognizing low resolution images and inadequate visual information, while the teacher's problem network includes a great number of characteristics. The student's network is condensed to a residual network with only three residual blocks, which is the breadth of the wide residual network, or the instructor's network, when the teacher chooses a network with an extensive residual network depth of 40. In order to promote the transfer of broken knowledge to the student network, the teacher network's capacity to represent the features of the broken images is first enhanced by extending the convolutional channel dimension. The student network is also simplified to a 10-layer residual network with three residual blocks in order to considerably minimize the number of student network parameters. The degree of fragmented knowledge transfer from the teacher network to the student network was visualized by calculating the ratio of the difference in correct rates between the teacher network of known width and the mentored student network to the difference *i*. Next, knowledge bias was measured by mapping the performance of teacher and student networks of different widths on a number axis. The degree of information transfer increases with decreasing knowledge divergence. Finally, the knowledge deviation and degree of accuracy improvement of the student network under various methods are combined to determine the best knowledge transfer network.

In order to guide the student network after training to obtain the best student network for broken classification accuracy, the teacher network is used in conjunction with a combination of hidden layer attention, algorithm, and output layer knowledge distillation algorithm[10] for maximum knowledge transfer.

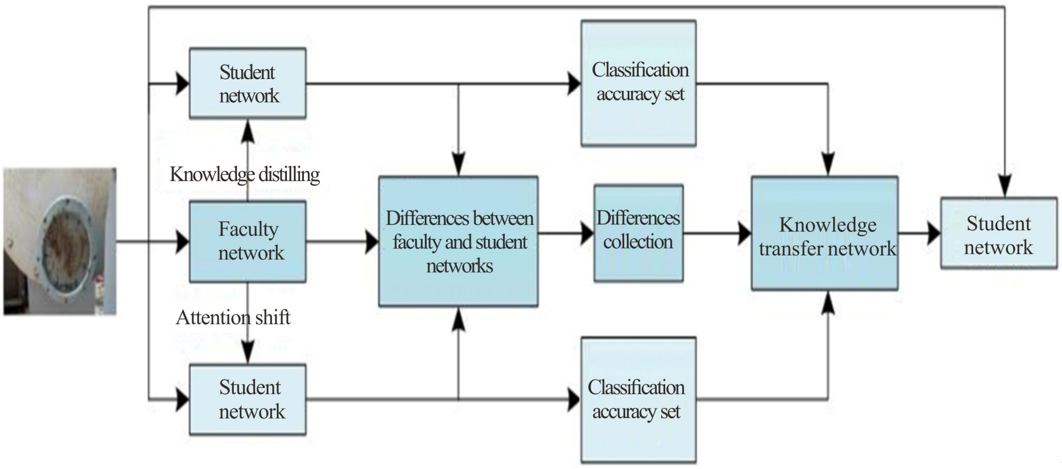

Fig.1 Broken image classification process of the teacher-student network

The complex imaging backgrounds and significant variations in distance and angle in the detection images are distinguishing features. Additionally, the instrument panel images typically have low resolution due to the low proportion of instrument panel photos in the detection images[11]. As a result, the instrument panel broken image classification dataset contains a significant amount of challenging samples, such as shadows, blurring, and occlusions, which significantly raises the challenge of the task. Because the dashboard image has low resolution and poor visual information, the feature information in the image is fully extracted by widening the feature expression dimension of the image, which increases the transferable defect knowledge from the teacher network to the student network and boosts the student network's classification accuracy. The number of convolutional kernels in each convolutional layer in the network is

$$M = K * N, N \in [16, 32, 64], \tag{7}$$

where 16, 32 and 64 are used to represent the bases of the network's lower, middle, and higher convolutional kernels. The dashboard defect image's feature information emphasis and the network's bottom, middle, and higher convolutional kernel counts vary.

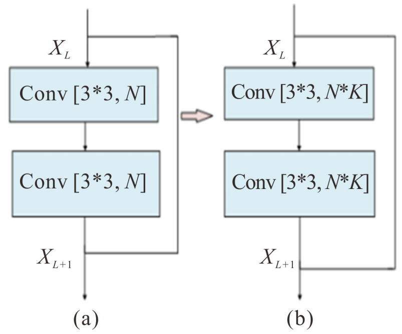

Fig.2 Residual blocks: (a) Regular residual block; (b) Widened residual block

The convolutional kernels of each convolutional layer are still parallel and can completely exploit the feature information in the dashboard image even after widening the network. The network width is not positively connected with the classification effect, which is predicted, and the width cannot be prolonged forever. This is due to the fact that the dashboard image contains additional redundant information in addition to the data information, and an excessive increase in network width will result in an overextraction of the features and other information present in the dashboard image, increasing the likelihood of network error learning and affecting the final student network classification results.

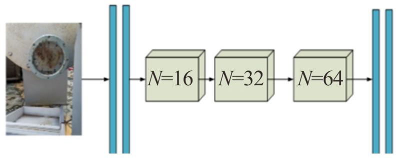

Fig.3 Student network structure

The student network was narrowed down to a 10-layer residual network with three residual blocks in order to considerably minimize the characteristics of the student network. The student network reserves 16, 32 and 64 convolutional kernels in the low, middle, and high layers of the network, respectively, to guarantee that it perfectly matches the teacher network's ability to represent features in the hidden layer. The student network is a fundamental residual network[12], and the goal of selecting three distinct convolutional kernel number residual blocks is to decrease the number.

4 000 transformer instrument panel images are gathered for this paper and divided into 8 categories based on the actual situation, of which 1 823 complete and 1 005 basically complete images are used as the majority category. The remaining 622 mirror blurred images are 338 mirror broken images, 85 mirror aging images, 97 mirror tinted images, 13 remaining 622 mirror blurred images, 338 mirror broken images, 85 mirror degraded images, and 97 dashboard stained images. The number of majority and minority classes in this dataset, which is sampled and trained using the same methodology, differs significantly. This dataset is known as Dashboard-V1. We choose 2 400 photos at random for training, 300 for validation, and the remaining 300 for testing in the experiment. The test set is divided into four input blocks of pixel size, and the output is stitched into pixels since the image size of the training set is pixels, whereas the image size of the validation set and the test set is pixels.

All models in this study were implemented and trained using PyTorch[13] GPU version, the Dashboard-V1 dataset models were trained on a single NVIDIA GTX 1080Ti with an 8-model batch size. The network was trained using stochastic gradient descent (SGD)[14] for 200 iterations (momentum). The learning rate is first set to 0.1 and then decreased by 0.01 after 160 and 180 iterations, respectively. Using the common cross-first loss, we train the network for the first 100 stages, and then we use the impact equilibrium loss to fine-tune the network for the following 100 stages. Our suggested technique achieves 84% accuracy in surface breakage and 86% accuracy in surface blurring on the Dashboard-V1 dataset.

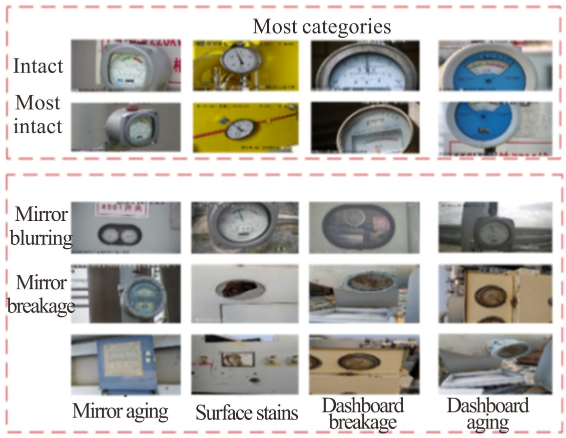

Fig.4 Dashboard-V1 partial dataset

Tab.1 displays the findings. The accuracy of the student network increased by 2.26% after the teacher network used a combination of implicit layer attention transfer and output layer knowledge distillation to direct the training of the student network. The accuracy gap with the teacher network was then only 2.82%, with a knowledge bias of 0.28, and the student network's participant count was only 0.56% of that of the teacher network.

Tab.1 Accuracy of teacher-student network classification

Residual network	Accuracy (%)	Number of participants	Number of bytes
ResNet-10	83.12	78 330	0.30
ResNet-10	85.78	78 330	0.30
ResNet-40	**88.56**	**132 965**	**132.86**

In conclusion, this work proposes a novel effect of balance loss in order to address the overfitting problem with the majority of classes in the class imbalance problem. The identification issue of broken substation instrument panels can be identified by the model trained on the unbalanced class data. Due to its direct focus on the impact of samples on the model, the impact balance loss described in this research can robustly assign weights, in contrast to prior methods. Additionally, a solution for the issue that the low resolution of the detected images and inadequate visual information leads to a large number of teacher network parameters and low accuracy of the student network is proposed by utilizing deep learning and edge computing techniques. With the development of edge computing technology in the future, the terminal equipment and edge infrastructure of the power system may have edge computing capability, and edge computing will have a broader space for development in the electric power intelligent network.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] QIU T, CHI J, ZHOU X, et al. Edge computing in industrial internet of things：architecture, advances and challenges[J]. IEEE communications surveys & tutorials, 2020, 22(4)：2462-2488.

[2] SHANG F, LAI J, CHEN J, et al. A model compression based framework for electrical equipment intelligent inspection on edge computing environment[C]//2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), April 24-26, 2021, Chengdu, China. New York：IEEE, 2021：406-410.

[3] ZHU X, YANG J, ZHANG C, et al. Efficient utilization of missing data in cost-sensitive learning[J]. IEEE transactions on knowledge and data engineering, 2019, 33(6)：2425-2436.

[4] WU J, BRAVERMAN V, YANG L. Obtaining adjustable regularization for free via iterate averaging[C]//International Conference on Machine Learning, July 12-18, 2020, Vienna, Austria. Princeton：IMLS, 2020：10344-10354.

[5] RIZK H, YAMAGUCHI H, HIGASHINO T, et al. A ubiquitous and accurate floor estimation system using deep representational learning[C]//Proceedings of the 28th International Conference on Advances in Geographic Information Systems, November 3-6, 2020, Seattle, WA, USA. New York：ACM, 2020：540-549.

[6] DENG S, ZHAO H, FANG W, et al. Edge intelligence：the confluence of edge computing and artificial intelligence[J]. IEEE internet of things journal, 2020, 7(8)：7457-7469.

[7] ZHOU Z, CHEN X, LI E, et al. Edge intelligence：paving the last mile of artificial intelligence with edge computing[J]. Proceedings of the IEEE, 2019, 107(8)：1738-1762.

[8] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. Digital signal processing, 2022：103514.

[9] LI L, DOROSLOVAČKI M, LOEW M H. Approximating the gradient of cross-entropy loss function[J]. IEEE access, 2020, 8：111626-111635.

[10] PARK D Y, CHA M H, KIM D, et al. Learning student-friendly teacher networks for knowledge distillation[J]. Advances in neural information processing systems, 2021, 34：13292-13303.

[11] CHU Y, YANG X, LI H, et al. Multi-level feature aggregation network for instrument identification of endoscopic images[J]. Physics in medicine & biology, 2020, 65(16)：165004.

[12] SINGADKAR G, MAHAJAN A, THAKUR M, et al. Deep deconvolutional residual network based automatic lung nodule segmentation[J]. Journal of digital imaging, 2020, 33(3)：678-684.

[13] PASZKE A, GROSS S, MASSA F, et al. Pytorch：an imperative style, high-performance deep learning library[EB/OL]. (2019-12-03) [2022-09-20]. https：//arxiv.org/abs/1912.01703.

[14] LI Y, LIANG Y. Learning overparameterized neural networks via stochastic gradient descent on structured data[EB/OL]. (2019-08-01) [2022-09-20]. https：//arxiv.org/abs/1808.01204v1.