ELEKTROTEHNIŠKI VESTNIK 86(5): 259–266, 2019
ORIGINAL SCIENTIFIC PAPER

Embedded Hoist-Intrusion Detection Using Deep Learning

Lizhong Zhang[1], Mei Li[1], Fei Guo[1], Ninghui He[1], Weiguo Sha[1], Bo Wang[2], Junling Zhang[3], Jiehan Zhou[4], Hongwei Zhao[5], Weishan Zhang[5]

[1]*State Grid Ningxia Electric Power Co.Ltd, Yinchuan, China*
[2]*Wuzhong Power Supply Co., Ningxia Electric Power Co.Ltd, Wuzhong, China*
[3]*Shandong Luneng Software Teconology Co.Ltd, Jinan, China*
[4]*Oulu University, Finland*
[5]*China University of Petroleum, Qingdao, China*
E-mail:zhangws@upc.edu.cn

Abstract. Hoist intrusion threatens the power transmission line safety. The paper proposes an embedded hoist intrusion detection approach using deep learning (EHIDuDL). A new hoist detection network (HDNet) is designed which contains a hoist rough-detection network (HRDNet) and a hoist classification-network (HCNet). HRDNet shares its basic convolutional layers with HCNet to reduce parameter setting and improve detection efficiency. Different data enhancement methods are used to expand the data set. The proposed model is compressed by channel pruning. A practical deployment and evaluation demonstrate the EHIDuDL reliability and efficiency.

Keywords: Deep learning; Hoist detection; HRDNet; HCNet

Vgrajeno zaznavanje vdora žerjava z uporabo globokega učenja

Žerjavi ogrožajo varnost električnih daljnovodov. V prispevku predstavljamo vgrajeno mrežo za zaznavanje vdora žerjava v območje daljnovoda z uporabo globokega učenja. Sistem je sestavljen iz mreže za grobo zaznavanje HRDNet in klasifikacijske mreže HCNet. Mreži si delita osnovne konvolucijske plasti, s čimer zmanjšamo število parametrov in učinkovitost zaznavanja. Nabor podatkov smo razširili z metodami izboljšanja podatkov. Predlagan model smo skrčili z rezanjem nepomembnih kanalov. Praktična uporaba in ocena delovanja dokazujeta zanesljivost in učinkovitost predlaganega pristopa.

1 INTRODUCTION

The safety of the power transmission lines is important. The safety of transmission lines will be threatened when these hoist lifting jibs are running in its vicinity. Therefore, it is particularly valuable to design a hoist intrusion detection system to prevent such threats.

There are several object detection methods that can be considered for hoist detection. [1], [2]. The single-stage methods, such as SSD [3], YOLO [4] and CornerNet [5], directly generate detection results from images. The two-stage methods such as R-CNN [6], Fast R-CNN [7], Faster R-CNN [8] and Mask R-CNN [9], first extract the candidate image regions, and their features from candidate regions are then used to get object classifications. Although these algorithms have a good performance with powerful machines, it is difficult to run them directly on an embedded system with a solar-cell-powered video camera to support full-day operation. There are lightweight solutions such as YOLO-Lite [10], MobileNet [11] and ShuffleNet [12]. Although they improve the image processing speed and efficiency, their low accuracy makes it difficult to meet the industrial requirement. Therefore, it needs a new method to detect the hoist intrusion for the purpose of power transmission line safety.

The paper proposes a deep-learning-based hoist intrusion detection called EHIDuDL. The system reduces the energy consumption by determining whether or not to turn the equipment on with infrared sensors [13]. A new hoist-detection network(HDNet) containing a Hoist Rough Detection Network (HRDNet) and a Hoist Classification Network (HCNet) is proposed. HRDNet is responsible for a preliminary target detection to provide candidate target areas and HCNet is responsible for the candidate areas classification. HCNet reduces computation by sharing some of the feature extraction network [14] with HRDNet. The traditional data enhancement methods [15] and GAN [16] are used to expand the data sets. The channel-pruning strategy [17] is used to compress the model [18] resulting in the reduced energy consumption, allowing it to run on a lightweight computing platform and improve the recognition speed. Experimental results demonstrate that EHIDuDL can accurately detect the hoist intrusion and has a high availability and stability.

Received 8 April 2019
Accepted 24 October 2019

The main contributions of the paper are:

- An efficient deep-learning-based hoist-intrusion detection approach is proposed for detecting and alarming hoist intrusions.
- HRDNet is proposed to detect hoist intrusion and to share the base convolutional layers with HCNet, thus reducing the parameters and improving the efficiency.
- An innovative idea using a smaller classification network for object detection is proposed, paying a particular attention to regional proposals.

The paper is organized as follows, Section 2 discusses the related work. Section 3 presents the EHIDuDL design and implementation. Section 4 evaluates the proposed solution. Section 5 concludes the paper.

2 RELATED WORK

Deep learning has gradually replaced the traditional detection algorithm and has become the mainstream method of the object detection since Hinton et al. [2] put forward the concept of deep learning in 2012. So far, it has made remarkable achievements in the field of image processing. R-CNN [6], which is the pioneer of using deep learning in object detection algorithm, is the representative of the two-step method. It first generates regions of interest, and then uses CNN to recognize and classify them. He Kaiming et al. [19] further optimize R-CNN and propose SPP-Net. In their algorithm, a spatial pyramid pooling layer is added between the convolutional layer and a fully connected layer of the network so that any-size feature map can be transformed into a fixed-size feature vector. Girshick [7] proposes Fast R-CNN which adopts SPP-Net on the basis of R-CNN and improves R-CNN to get a better performance. The key of Faster R-CNN lies in the design of RPN (Regional Proposal Network) which generates better region proposals. In 2017, He Kaiming et al. [9] continue to improve R-CNN on the basis of the previous one and propose Mask R-CNN which upgrades the RoI Pooling layer of Fast R-CNN to the RoI Align layer. They also add a new branch to handle the semantic segmentation.

As to the one-stage methods, YOLO [4] performs a real-time detection of an end-to-end deep learning system. The main idea of the YOLO series is to get the classes and the specific position of the object directly from the input image instead of generating region proposals like the R-CNN series. A significant feature of YOLO is its efficiency.

The object detection on embedded devices has been attracting a lot of attention. In [20], a research performs Fast R-CNN on Jetson TK1. Although the researchers make some additional modifications on Fast R-CNN, the inference efficiency was still very low (1.85 fps). In order to reduce the inference time on embedded devices, two lightweight networks, MobileNet [11] and ShuffleNet [12], are proposed recently. However, the decrease in the model volume leads to the decrease in the detection accuracy. Therefore, a powerful object-detection method is needed to support the real-time image processing on embedded devices.

3 EHIDUDL DESIGN AND IMPLEMENTATION

The EHIDuDL goal is to achieve a good detection accuracy and inference efficiency. To reduce the energy consumption, solar-powered infrared sensors are used to control operation. When an infrared sensor detects some nearby machines, it will wake up the camera and also the hoist detection. After decoding the captured video, the images will be transmitted to the hoist-detection network. The hoist detection network precisely proposes candidate regions and the smaller classification network simplifies the network without accuracy loss. To further reduce the number of its parameters, the model compression techniques are designed. The detection results are sent to a back-end server to trigger alarms. Figure 1 presents the overall EHIDuDL architecture.

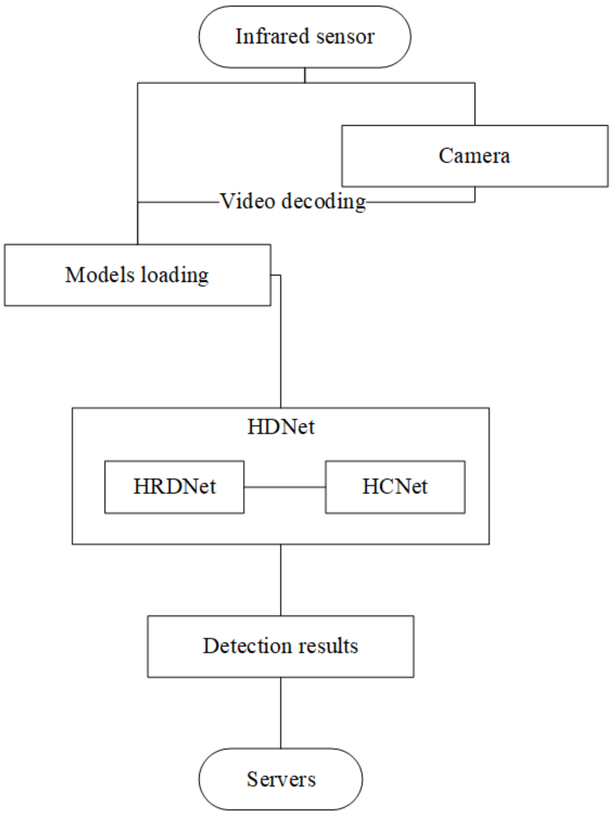

Figure 1. Overall EHIDuDL architecture

In the following, the enhancement of the data set for training an effective neural network will be discussed and the design of the neural network components will be presented.

3.1 Data enhancement

By combining the traditional image geometric transformation and GAN (Generative Adversarial Networks), the data set is expanded to address the problem of a small data set. Horizontal flipping, image tilting, noise adding and image scaling are used to expand the data set. Figure 2 shows an image with Gaussian noise.

Figure 2. Image generated with the traditional data enhancement method

DCGAN [21] is used to generate new images. Each image is resized to 960×640 and each batch contains 16 images considering the GPU memory limitation. The losses of the generator, discriminator and image samples generated by GAN are shown in Figures 3, 4 and 5.

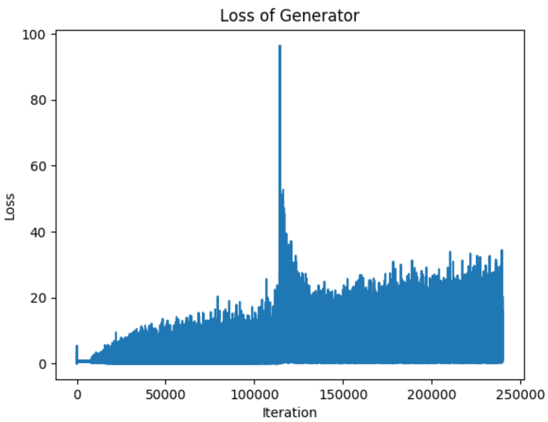

Figure 3. Generator

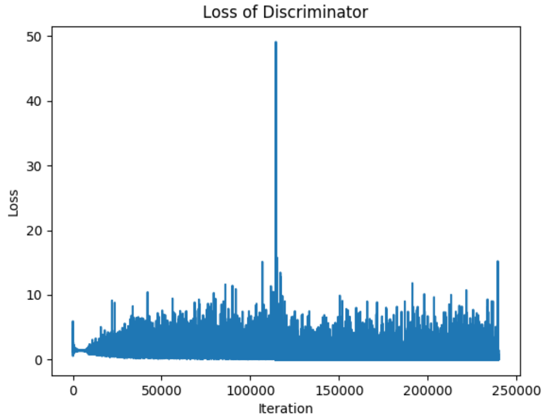

Figure 4. Discriminator loss

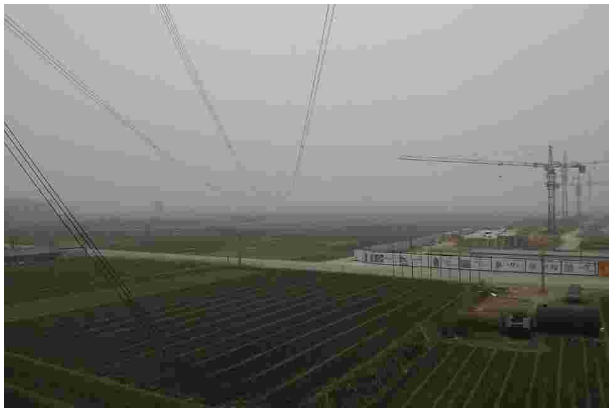

Figure 5. Image generated with GAN

3.2 HDNet

A new end-to-end hoisting machinery intrusion detection network HDNet is proposed. It is mainly composed of two parts: Hoist Rough-Detection Network (HRDNet) and Hoist Classification Network (HCNet). HDNet uses the model compression to reduce the number of its unnecessary parameters, compress the model volume and improve the inference speed. HRDNet is responsible for a rough detection of the candidate hoist areas. HRNet is responsible for further classification. HRDNet can be regarded as a precise Region Proposal Network(RPN) [8]. HCNet shares part of the HRDNet convolutional layer to reduce detection computation. Figure 6 presents the network structure.

3.2.1 HRDNet: HRDNet is motivated by the grid cells regression calculation in YOLO. The main model network is simplified as shown in Figure 7. At the same time, HRDNet does not need a separate classifier because the only target is to detect the hoist. So it abandons the branch of the class prediction and uses the confidence to measure the probability of the object in an anchor box. According to the hoist size in our data set, the size of the hoists are not fluctuating that much, so the feature pyramid structure [22] is abandoned to reduce computation overhead. Experiments show that the model accuracy is the highest when the image is divided into 9×9 grids.

3.2.2 HCNet: HCNet shares a part of the convolutional layers of HRDNet as shown in Figure 6. HCNet

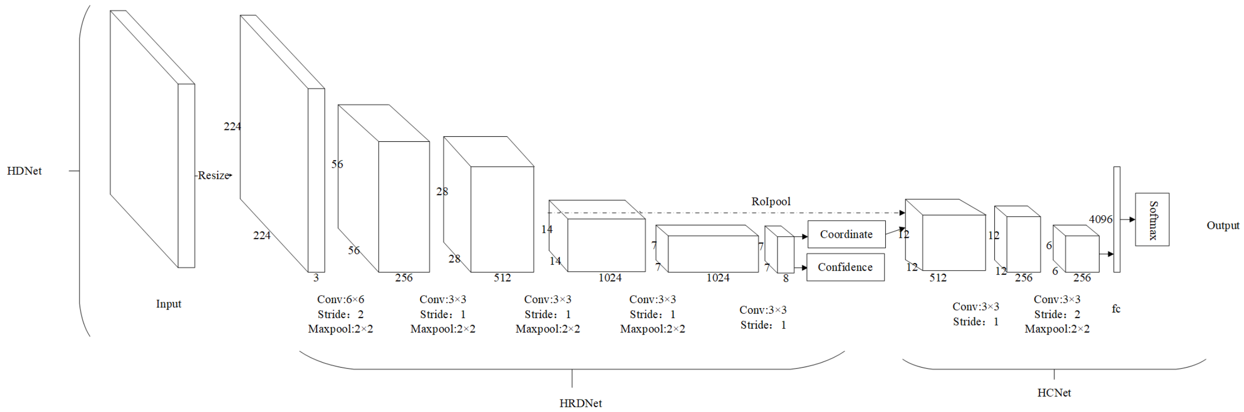

Figure 6. HDNet

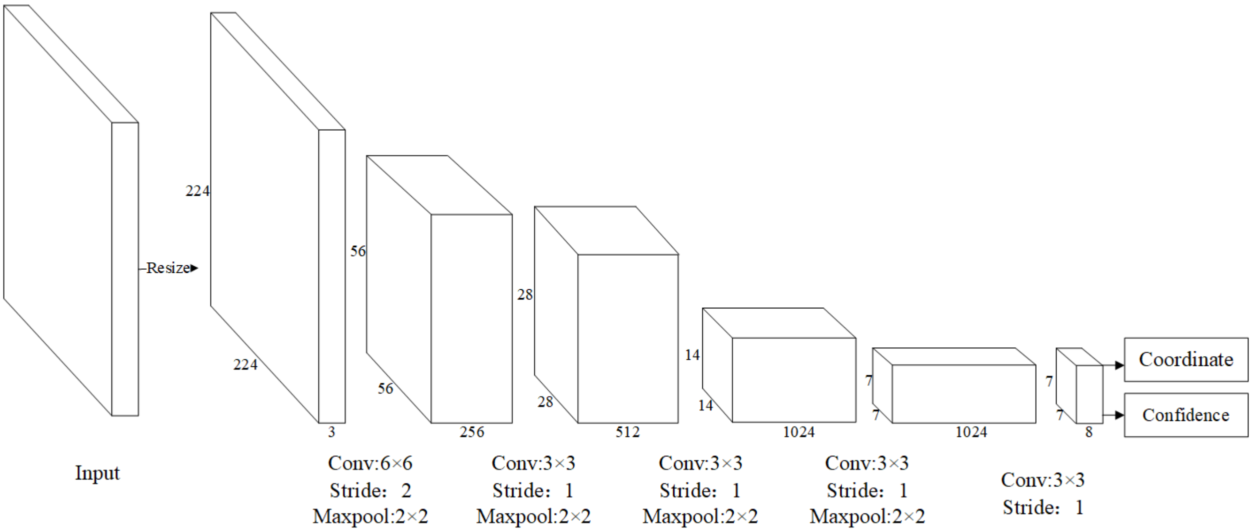

Figure 7. HRDNet structure

does not need to extract features from the original image to classify the candidate regions. Therefore, the number of the layers of this classification network does not need to be calculated, which greatly reduces the calculation overhead brought by the image feature extraction. The region coordinates obtained by the HRDNet regression calculation, they are mapped to the HRDNet feature map obtained by the convolution operation. HCNet cuts out the candidate regions according to the region coordinates on the feature map. These regions of interest will be resized to a uniform size of 12×12×512 by RoI Pooling [8] and bilinear interpolation. A vector of the length of 4096 is generated through two convolutional layers and one full-connection layer [23]. According to the vector, its classification is obtained with Softmax to see whether the candidate region is a hoist. The HCNet architecture is shown in Figure 8.

3.2.3 Loss Function: Since HDNet only needs to detect one class instead of performing a multi-category classification, the final classification loss function only needs to judge the confidence of a single category. The confidence loss is expressed as a cross-entropy loss function [25] as follows:

$$L_{Confidence} = \sum_{i=0}^{S^2} [C_i log C_i^* + (1-C_i)(1-log C_i^*)] \quad (1)$$

S^2 is the total number of grid cells, C_i is the confidence of the predicted results generated by HDNet, and C_i^* is the truth. If there are no objects in the predictions, then $C_i^* = 0$.

The loss function of the coordinate is shown in Equation 2 where (x_i, y_i) is the center coordinate of the prediction result and (x_i^*, y_i^*) is the true center coordinate. Correspondingly, (w_i, h_i) and (w_i^*, h_i^*) are the width and height of the predicted and real targets respectively.

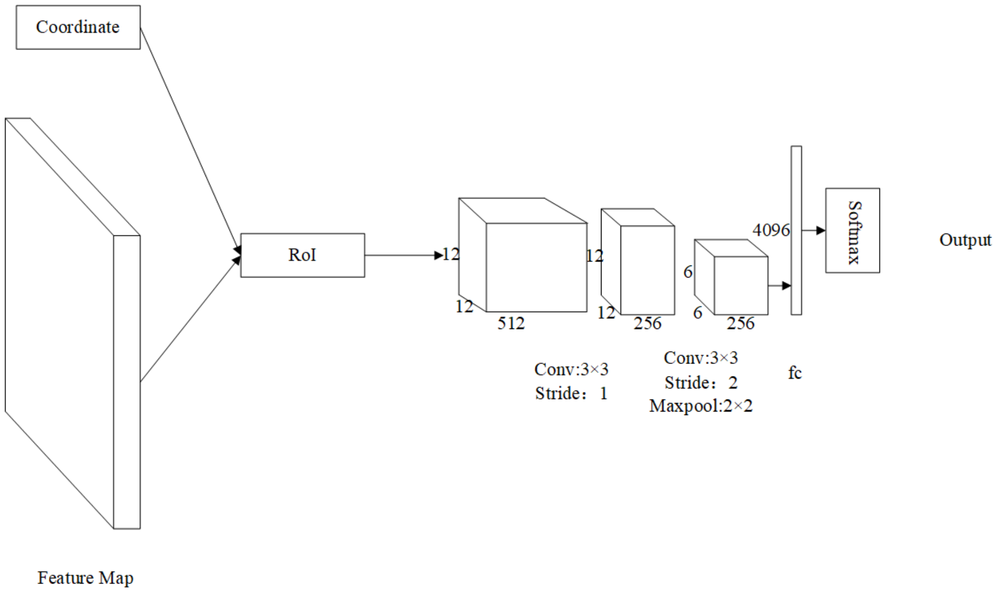

Figure 8. HCNet structure

$$
\begin{aligned}
L_{Coordinate} = \sum_{i \in propers} & [x_i log x_i^* + (1 - x_i)(1 - log x_i^*) \\
& + y_i log y_i^* + (1 - y_i)(1 - log y_i^*) \\
& + \frac{\sqrt{w_i} log w_i^*}{2} \\
& + (1 - \sqrt{w_i})(1 - \frac{log w_i^*}{2}) \\
& + \frac{\sqrt{h_i} log h_i^*}{2} \\
& + (1 - \sqrt{h_i})(1 - \frac{log h_i^*}{2})]
\end{aligned} \tag{2}
$$

Using the flow chart of HDNet shown in Figure 9, the HDNet loss function is designed as a weighted sum of confidence loss and coordinate loss as follows.

$$Loss = \lambda L_{Confidence} + (1 - \lambda) L_{Coordinate} \tag{3}$$

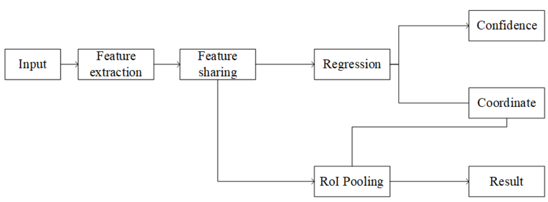

Figure 9. HDNet flow chart

3.2.4 Model compression: The model is compressed by channel pruning. For the trained model, the unimportant channels are pruned meaning that these channels are deleted and a new model map is constructed. The model is then retrained to restore the accuracy.

For each filter in the channel level, its Frobenius norm is calculated and then binaryzation is applied to it. If its Frobenius norm is greater than 0, it is changed into 1. Correspondingly, if it is equal to 0, it is not changed. The Frobenius norm of a channel is summed to find the redundant channel in each layer of the neural network. Then the upstream neurons of the redundant channels can be deleted safely. This can also reduce the number of the downstream channels from the deleted neurons, so as to prune the branches which have less effect on the neural network to obtain a smaller model. The model obtained by pruning is retrained with the training set to make up for the accuracy loss caused by pruning. In this way, the volume of the model is compressed and the inference speed is improved without loss of model accuracy.

4 EXPERIMENT AND ANALYSIS

In order to test the HDNet performance, our experiments are carried out using NVIDIA TX2. The software environment is shown in Table 1.

The accuracy and recall are the standard to evaluate the performance of the algorithm and model.

$$Precision = \frac{TP}{YP + FP} \tag{4}$$

$$Recall = \frac{TP}{YP + FN} \tag{5}$$

TP is the number of the correct detection results, FN is the number of the missed results, and FP is the number of the false results. To warn the invading hoist, we only need to judge whether there is a hoist

Table 1. Software environment

Software	Version
Python	3.5
JDK	1.8
Pytorch	1.0
Numpy	3.2
ANT	1.9.7
Scipy	1.0.1
OpenCV	3.2
Darknet	1.0

in the image and the coordinates accuracy is not so important. So when the Intersection over Union(IoU) of the detection coordinate results and the real value is greater than 0.25, the detection is considered correct .

4.1 Grid cells

Considering the effect of the size of hoist place in the image on the detection results, the detection models are tested with the grid cells under different scales in order to select the best scale of the HRDNet grid cells. The test results are shown in the Table 2.

Table 2. Comparison of different division strategies

	Precision	Recall	Efficiency
2×2	0.8023	0.6759	33.61
3×3	0.8530	0.7199	36.48
4×4	0.8605	0.8623	40.84
5×5	0.8830	0.8690	42.62
6×6	0.9491	0.8534	42.14
7×7	0.9176	0.8865	44.32
8×8	0.9262	0.9014	47.66
9×9	0.9429	0.9205	50.53
10×10	0.9248	0.9155	67.21
11×11	0.9051	0.8759	95.30
12×12	0.8214	0.8603	130.96
13×13	0.8445	0.8821	186.51
14×14	0.8061	0.8449	220.62
15×15	0.7251	0.8650	248.12

Our experimental results show that with the number of the grids increasing, the processing time of the image increases, too. If the size of the image grids is too large (2×2), the recall is low and there is a missed detection. When the size of the image grids is too small (15×15), the accuracy of the model is low, and a false detection will occur. The accuracy and recall rate of the model are the highest when the size of the image grids is 9×9. Therefore, 9×9 is selected as the best image divisioning.

4.2 Data enhancement

In order to improve the effectiveness of a neural network and the contribution of the data enhancement to the accuracy, the collected data set is enlarged with both the classical approach and the GAN-based approach. 1000 original images are prepared, 2000 new images are generated with a traditional image geometry transformation (GT) and 2000 new images are generated from the original ones with GAN. The experimental results are shown in Table 3. The conclusion from our experimental results is that a combination of the image transformation and GAN can significantly improve the precision and recall rate of the model.

Table 3. Detection results with the data enhancement

	Precision	Recall
Original Image	0.8151	0.6381
GT	0.8362	0.7151
GAN	0.9396	0.9015
GT+GAN	0.9429	0.9205

4.3 Input scales

Generally speaking, the higher the resolution of an input image is, the higher accuracy of a neural network can detect, and moreover, the inference time will also increases. In order to balance the model accuracy with the efficiency of inference, different resolutions of an input image are tested. The test results are shown in Table 4.

Table 4. Results of different input scales

	Precision	Recall	Efficiency
240×240	0.5854	0.4216	39.15
320×320	0.6163	0.6048	41.35
416×416	0.7761	0.8186	45.69
480×480	0.9429	0.9205	50.53
640×640	0.9160	0.9415	195.18
960×960	0.9561	0.9784	318.62

To meet the need of practical application scenarios, the 480×480 input resolution are choosen.

Table 5. Results of different input scales

	Precision	Recall	Efficiency
HDNet	0.9429	0.9205	50.53
MobileNet	0.7835	0.8447	68.48
ShuffleNet	0.8165	0.8491	53.98
HDNet without compression	0.9368	0.9315	243.66

4.4 Comparison with other methods

In order to test the HDNet performance, it is compared with uncompressed HDNet, MobileNet and ShuffleNet for both the accuracy and efficiency. The experimental results are shown in Table 5. As seen the HDNet accuracy is close to the accuracy of the uncompressed HDNet, and its efficiency is much higher than that of the uncompressed HDNet. Compared with other lightweight target detection models, HDNet outperforms them both in terms of the accuracy and efficiency.

4.5 Industrial application

The proposed EHIDuDL is deployed to the State Grid in Ningxia. The camera used is HIKVISION DS-2CD3T25D-I5, installed at the height of 25-40 meters on 50 transmission towers. The image resolution is 1920×1080 and the computing platform is NVIDIA Tegra X2. The results of a practical detection are shown in Figure 10.

Figure 10. Detection result

In the last nine months of our deployment, 63 alarms were collected, 61 hoist intrusion occurred among which 59 were correctly detected. The EHIDuDL detection records are shown in Table 6.

Table 6. Results of different input scales

	Alarm	Truth	Precision	Recall
EHIDuDL	63	59	0.9365	0.9672

EHIDuDL correctly detects the hoist intrusion in the power transmission line scenario with high recall rate. Its reliable and efficient running demonstrates that the proposed EHIDuDL is performing well.

5 Conclusion

The paper presents a deep-learning-based hoist-intrusion detection approach (EHIDuDL) for power transmission line scenarios. A new hoist detection network(HDNet), hoist classification network(HCNet) and model compression technique are designed. The HRDNet sharing the basic convolutional layers with HCNet reduces the parameter setting and improves the detection efficiency. To compress the proposed model, channel pruning is designed. Experimental results and practical deployment show that EHIDuDL works reliably, effectively and efficiently.

Acknowledgement

This research was supported in part by the National Major Science and Technology Project (2017ZX05013-002), the Key Research Program of Shandong Province under Grant 2017GGX10140 and in part by the Fundamental Research Funds for the Central Universities

References

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436.
[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[3] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.
[4] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[5] Law, Hei, and Jia Deng. "Cornernet: Detecting objects as paired keypoints." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
[6] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
[7] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.
[8] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.
[9] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[10] Pedoeem, Jonathan, and Rachel Huang. ”Yolo-lite: A real-time object detection algorithm optimized for non-gpu computers.” arXiv preprint arXiv:1811.05588 (2018).
[11] Howard, Andrew G., et al. ”Mobilenets: Efficient convolutional neural networks for mobile vision applications.” arXiv preprint arXiv:1704.04861 (2017).
[12] Hluchyj, Michael G., and Mark J. Karol. ”Shuffle Net: An application of generalized perfect shuffles to multihop lightwave networks.” Journal of Lightwave Technology 9.10 (1991): 1386-1397.
[13] Jha, A. R. ”Infrared technology, applications to electro-optics, photonic devices, and sensors.” Infrared 11 (2000): 1944-1948.
[14] Lu, Yongxi, et al. ”Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
[15] Lamburt, Leonid, and Lazar Koyfman. ”Data enhancement techniques.” U.S. Patent No. 6,397,228. 28 May 2002.
[16] Goodfellow, Ian, et al. ”Generative adversarial nets.” Advances in neural information processing systems. 2014.
[17] He, Yihui, Xiangyu Zhang, and Jian Sun. ”Channel pruning for accelerating very deep neural networks.” Proceedings of the IEEE International Conference on Computer Vision. 2017.
[18] Han, Song, Huizi Mao, and William J. Dally. ”Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding.” arXiv preprint arXiv:1510.00149 (2015).
[19] He, Kaiming, et al. ”Spatial pyramid pooling in deep convolutional networks for visual recognition.” IEEE transactions on pattern analysis and machine intelligence 37.9 (2015): 1904-1916.
[20] Jagannathan, Shyam, et al. ”Efficient object detection and classification on low power embedded systems.” 2017 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2017.
[21] Radford, Alec, Luke Metz, and Soumith Chintala. ”Unsupervised representation learning with deep convolutional generative adversarial networks.” arXiv preprint arXiv:1511.06434 (2015).
[22] Lin, Tsung-Yi, et al. ”Feature pyramid networks for object detection.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
[23] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. ”Fully convolutional networks for semantic segmentation.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[24] Liu, Weiyang, et al. ”Large-Margin Softmax Loss for Convolutional Neural Networks.” ICML. Vol. 2. No. 3. 2016.
[25] Schmidhuber, Jurgen. ”Deep learning in neural networks: An overview.” Neural networks 61 (2015): 85-117.

Lizhong Zhang Lizhong Zhang is a senior engineer at the Ningxia Electric Power Co.Ltd, China. The focus of his work on big data processing, power grid operation monitoring and maintenance.

Bo Wang Bo Wang is a senior engineer at the Wuzhong Power Supply Company of the Ningxia Electric Power Co.Ltd, China. The focus of his work is the big data analysis in electric power system operation monitoring.

Mei Li Mei Li is a senior engineer at the Ningxia Electric Power Co.Ltd. In 1993, She graduated from the Xi'an Jiaotong University. Her main research interests are mobile application and big data analysis technologies supporting power grid operation.

Weiguo Sha Weiguo Sha is a senior engineer at the Ningxia Electric Power Co.Ltd, China. His research focuses on big data analysis technology in electric power system operation monitoring.

Ninghui He Ninghui is a senior engineer at the Electric Power research institute of State Grid Ningxia Electric Power Co.Ltd, China. In 2013, he graduated from the Wuhan University of Technology and has a Ph.D. degree in automation. The focus of his work is on electrical equipment condition monitoring.

Jiehan Zhou Jiehan Zhou is currently an associate professor of the University of Oulu, Finland. He has acchieved his PhD degrees in computer engineering from the University of Oulu and in Manufacturing Automation from the Huazhong University of Science and Technology, China. His current research interests are in big data platforms, Internet of Things and pervasive cloud computing, He has published over 100 papers.

Fei Guo Fei Guo is a research engineer at the Ningxia Electric Power Co.Ltd, China. In 2007, he graduated from the Zhengzhou University with master degree in electrical engineering and automation. His current research interest is in big data analysis for power grid operations.

Junling Zhang Junling Zhang is a research engineer at the Shandong Luneng Software Teconology Co.Ltd, China. The focus of his work is on AI and big data analysis in electric power industry.

Hongwei Zhao Hongwei Zhao is a master student of the China University of Petroleum. The focus of his work is on deep learning and computer vision.

Weishan Zhang Weishan Zhang is a full professor at the Department of Software Engineering of the China University of Petroleum. In 2001, he got his PhD degree from the Northwestern Polytechnical University, China. His current research interests are in big data platforms, pervasive cloud computing and service-oriented computing. He has published over 120 papers and his current H-index according to the Google scholar is 17.