INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Advanced Mechatronic Systems

ISSN online: 1756-8420 - ISSN print: 1756-8412
https://www.inderscience.com/ijamechs

TTower-345: a multi-categories multi-perspectives benchmark for automatic naming of transmission line inspection photos

Jinqing Shen, Hong Ye, Chunjun Tang, Guoqin Zhang, Yan He, Min Xie

DOI: 10.1504/IJAMECHS.2024.10063115

Article History:

Received:	13 July 2023
Last revised:	20 September 2023
Accepted:	23 October 2023
Published online:	25 March 2024

Copyright © 2024 Inderscience Enterprises Ltd.

TTower-345: a multi-categories multi-perspectives benchmark for automatic naming of transmission line inspection photos

Jinqing Shen and Hong Ye

Jinhua Bada Group Ltd.,
Jinhua, Zhejiang, China
Email: Shenjq15685@sina.cn
Email: 380870918@qq.com

Chunjun Tang

Jinhua Power Supply Company of State Grid Zhejiang Electric Power Co.,
Jinhua, Zhejiang, China
Email: 577772@qq.com

Guoqin Zhang

China Jiliang Uinversity,
Hangzhou, China
Email: 908206887@qq.com

Yan He

Jinhua Power Supply Company of State Grid Zhejiang Electric Power Co.,
Jinhua, Zhejiang, China
Email: 1507609@qq.com

Min Xie*

China Jiliang University,
Hangzhou, China
Email: zjxiemin@163.com
*Corresponding author

Abstract: Efficient naming of inspection photos of transmission line towers is vital in the maintenance of power grid equipment. Current inspection photo naming methods are mainly manual, which is neither rapid nor effective. Research on inspection photo naming is limited due to a shortage of inspection image datasets and low image resolution. Hence, we gathered inspection photos of real tangent towers using drones and created an inspection photo dataset TTower-345 for automatic naming model training purposes. We proposed an automatic naming model, IELC (improved EfficientNet network and LBP classification model), based on this dataset. IELC comprises a dual-branch structure that integrates a jointly improved EfficientNet model and an local binary patterns (LBP) classification model. Experimental results indicated that the proposed dataset contains more diverse inspection image features, which in turn helped the model learn more features. In our experiments, our proposed automatic naming method achieved a classification accuracy of over 95% and demonstrated reliability by exhibiting good generalisability in practical scenarios.

Keywords: transmission lines; tangent towers; benchmark; inspection photo naming; EfficientNet model; local binary pattern; LBP classification.

Reference to this paper should be made as follows: Shen, J., Ye, H., Tang, C., Zhang, G., He, Y. and Xie, M. (2024) 'TTower-345: a multi-categories multi-perspectives benchmark for automatic naming of transmission line inspection photos', *Int. J. Advanced Mechatronic Systems*, Vol. 11, No. 1, pp.11–25.

Biographical notes: Jinqing Shen received his Bachelor's in Electrical Engineering and Automation from the Shanghai Electric Power University in 2007. Since 2021, served as the Chairman and Party Secretary of Jinhua Bada Collective Co., Ltd. His research directions are power system and its automation. He also received an awards that 'Active Control Technology and Application of High Permeability Distributed Photovoltaic Distribution Network' won the first prize of Zhejiang Provincial Electric Power Co., Ltd. Science and Technology Progress

Copyright © 2024 Inderscience Enterprises Ltd.

Award, the second prize of Zhejiang Provincial Science and Technology Progress Award, and the second prize of State Grid Corporation of China Science and Technology Progress Award. He also received the ‘Distribution Network Factory Installation and Distribution Application’ project won the top ten transformation and application achievements of State Grid Zhejiang Electric Power Co., Ltd.

Hong Ye received a Bachelor’s in Engineering Mechanics from the Tongji University in 2002. He also obtained his Master’s in Engineering Mechanics from the Tongji University in 2005. He have been working at State Grid Jinhua Power Supply Company since May 2005, successively serving as a secretary of the Youth League Committee of State Grid Jinhua Company, the Secretary of the Party Branch of the Transmission and Inspection Office, and a manager of Bada Technology Information Branch. He has research direction digitisation of power systems.

Chunjun Tang obtained his Master’s in Industrial Engineering from the North China Electric Power University in 2022. From 2021 to present, served as the Director of the Transmission and Inspection Center (Live Working Center) of State Grid Jinhua Power Supply Company. His research directions include transmission line construction, construction, operation, live line maintenance, and lean operation and inspection management. He received his awards in ‘Development of Key Technologies and Equipment for Intelligent Protection of Live Working on Ultra High Voltage Transmission Lines’ won the second prize of the Science and Technology Progress Award from State Grid Zhejiang Electric Power Co., Ltd. in 2021.

Guoqin Zhang received his Bachelor’s degree from the China University of Metrology in 2021. From 2021 to present, pursuing a Master’s degree at the China University of Metrology. His research direction is electronic information.

Yan He obtained his Bachelor’s degree from the Wuhan University of Water Resources and Electric Power in 1999. He also obtained the Master of Engineering from the Zhejiang University in 2012. Since 1999, he have been serving as a deputy manager of Jinhua Bada Group Technology Information Branch at State Grid Jinhua Power Supply Company. His research directions power system and its automation.

Min Xie obtained her Master’s in Testing and Metrology Technology and Instruments from the Chinese Academy of Metrology. She has been working as an automation teacher at the School of Mechanical and Electrical Engineering, China University of Metrology since July 2002. His research direction are temperature measurement calibration, drone control and application. He has also won awards in 2014, ‘Research and development of online calibration technology and devices for thermometers and measuring instruments’, and second prize of the Science and Technology Innovation Inspection Award of the General Administration of Quality Supervision, Inspection and Quarantine of China.

1 Introduction

Efficient and precise naming of inspection photos is crucial for the detection of defects in transmission lines. These named photos serve multiple purposes, such as detecting insulators, spotting equipment defects, and facilitating autonomous navigation and inspection with drones (Tang et al., 2021). The process of naming inspection photos can be viewed as an image classification task, wherein the goal is to categorise inspection photos according to their associated tower locations.

High-voltage transmission lines traverse diverse geographical environments, including mountainous terrains, basins, and reservoir areas. Factors related to both structural aspects and the environment contribute to an increased likelihood of defects, encompassing issues such as insulator defects, shock absorber malfunctions, broken lighting poles, as well as the breakage and corrosion of towers, among other concerns. These faults have the potential to seriously jeopardise the safe operation of the power system, underscoring the need for meticulously devised inspection strategies aimed at their identification and rectification. Currently, the predominant method employed by electric power grids for inspections is the use of unmanned aerial vehicle (UAV) patrols (Xiren et al., 2020). Nevertheless, the development of intelligent processing techniques for power grid image data remains in a nascent stage, primarily owing to the ensuing challenges and limitations:

- Systematic inspection of photo datasets related to transmission lines is notably deficient. In the realm of inspection datasets, the predominant offerings include those focused on distribution network power equipment (Kehui et al., 2019) and tower insulators (Panigrahy and Karmakar, 2022), while comprehensive, publicly accessible transmission line datasets are notably scarce due to cost constraints and privacy considerations.
- There exists a deficiency in algorithms capable of automatically assigning names to tower inspection

photos. Maintenance personnel continue to rely on manual analysis and interpretation of grid inspection photos to ascertain the operational status and health of the line. This process is pivotal in shaping subsequent control measures and inspection strategies (Mn et al., 2022). Furthermore, the precision of manually assigned names varies from individual to individual, posing challenges in guaranteeing the accuracy of the naming process.

In this paper, we introduce a multi-location inspection photo dataset named TTower-345, which is derived from the aforementioned context. We have taken measures to ensure that the image capture process for this dataset strictly complies with the technical standards and inspection criteria governing transmission line equipment inspection. Furthermore, it is essential to note that all images contained in this dataset were acquired in real-world settings through the use of UAVs. This dataset exhibits two core attributes:

- The dataset comprises a wide array of diverse inspection images from 13 UAV inspection sites located on tangent towers. These images encompass a range of perspectives, including high-to-low and left-to-right angles.
- The dataset encompasses a total of 345 tangent towers and 7,516 inspection photos. Due to the existence of multiple identical photos taken at the same location, each tangent tower contains an average of approximately 13.96 images after initial screening. Detailed information about this dataset will be presented in Section 3 of this paper. It is worth noting that this dataset holds significant potential for training automatic naming models tailored to inspection photos.

In our work, we conducted a series of experiments involving training and testing of various deep learning networks using the TTower-345 dataset. To enhance the efficiency of the baseline model for this inspection photo dataset, we introduced selective convolutional descriptor aggregation (SCDA) technology, which is aimed at improving the structure of EfficientNet. Subsequently, we parallelised the enhanced EfficientNet network with LBP classification branches to achieve the optimal baseline model. The effectiveness of this model was validated by ablation experiments, and its performance was compared with transfer learning models.

Our research findings demonstrate that the TTower-345 inspection photo dataset effectively facilitates the learning of high-level transmission line features by deep network models, even when they exhibit different foregrounds but share the same background. By integrating the improved EfficientNet baseline model with the LBP classification branch, we achieved significantly higher performance indicators, resulting in a precise and efficient automatic naming model for inspection photos.

Our work makes the following primary contributions:

- We construct a novel inspection photo dataset, TTower-345, which showcases various perspectives of tangent towers. These images were captured through UAV photography during actual inspections of transmission lines, offering a diverse range of scenes and heights. Importantly, the dataset reflects real-world conditions, including lighting variations and terrain obstacles, ensuring its authenticity and practicality.
- In this paper, we introduce a lightweight EfficientNet baseline model for the classification of inspection photos. Building upon the comprehensive evaluation results of the inspection photo dataset, we develop the improved EfficientNet network and LBP classification model (IELC). The IELC model leverages the SCDA structure to enhance the EfficientNet_b0 model and employs a dual-branch architecture to connect the improved EfficientNet model with the LBP classification model. The innovative intelligent classification model proposed in this paper offers electric utility companies an automated solution for the inspection photo naming process.

2 Related work

2.1 Naming techniques for inspection photos

The field of power transmission line management has traditionally relied on manual naming patterns for photos, a process that consumes significant time and labor in image organisation and analysis. This manual approach is often hindered by variations in judgment standards among inspection and maintenance personnel, resulting in inconsistent analysis outcomes.

The emergence and widespread adoption of deep learning models in power transmission lines, as noted in prior research (Tianjiao et al., 2020; Gao et al., 2022), have paved the way for increased automation. Nevertheless, the domain of tower inspection photo naming has seen limited exploration. Most current research has focused on semi-automatic naming systems. For instance, Su's (2020) work introduced an efficient image standardisation and organisation tool for power transmission line photos, replacing manual typing with mouse clicks or dedicated console operations. Similarly, Guowei et al. (2020) proposed an automatic or semi-automatic approach to collect information related to the relationship between position of service (POS) data and tower spatial information to enhance post-data processing efficiency for machine inspection photos. Furthermore, Du et al. (2022) suggested a preprocessing approach rooted in drone inspection techniques and their applications, utilising human diagnostic technology to assist in assessing the inspection image location. However, these methods rely heavily on pre-training comparison models and exhibit large subjectivity, limiting their practical applicability.

In light of these challenges and the presence of specific inspection photos for each inspection position in our dataset, we devised a foundational model based on the EfficientNet network to classify these images. During the testing phase, our trained model categorises the inspection photos and autonomously assigns a name to each one. This novel approach significantly enhances data processing efficiency.

2.2 Dataset related to power transmission line

The majority of inspection photos primarily focus on capturing key sites of power transmission lines, primarily for defect detection purposes (Zhang et al., 2022) and image generation tasks. These photos serve as the foundation for creating various datasets, such as InsuGenSet (Zhao, 2021b) and Insulator Dataset (ID) (Xia et al., 2022). These datasets were developed in response to the limited image resolution of datasets like ImageNet or PascalVOC. They have been instrumental in training deep learning models with high accuracy and strong generalisation capabilities.

Furthermore, a recent initiative, the Key Component Image Generation Dataset (KCIGD), was launched to train an insulator image generation network. This effort aimed to address the challenge of plain backgrounds in images from the China Power Line Insulator Dataset (CPLID) (Tao et al., 2018). However, it is worth noting that the KCIGD (Wang et al., 2023) does not comprehensively represent the intricate backgrounds and authentic environments found in power transmission line inspection photos. The primary distinction between the former two datasets and the latter two lies in their objectives. The former datasets focus primarily on enhancing the resolution and authenticity of insulator images, while the latter ones aim to tackle the issue of plain backgrounds in the CPLID dataset and combine insulator images into photos. However, they do not fully capture the actual surroundings.

Consequently, existing datasets are typically limited to single inspection sites and lack the multi-perspective, multi-part characteristic information needed for full representation. Our proposed dataset addresses this limitation by including multiple inspection sites of the tower. It not only comprises insulator images but also provides diverse, real background information. This comprehensive dataset is instrumental in the development of automatic naming methods and defect detection networks for inspection photos.

3 TTower-345 dataset

3.1 Collection of datasets

Presently, many studies on inspection photos of transmission line towers rely on datasets like CPLID, InsuGenSet, ID, and others for insulator defect detection. However, these datasets are primarily suitable for specific inspection equipment classification and target detection, often lacking representation of a genuine environment. Therefore, in our work, we constructed a custom dataset using inspection photos. This dataset comprises a total of 7,516 images, each with dimensions of 5,472 $\times$ 3,078 pixels.

These images were acquired by a company during aerial patrolling of multiple 500 kV overhead transmission lines using drones. Notably, images captured from various inspection positions exhibit distinct shooting perspectives, and the image backgrounds are characterised by complexity, encompassing a variety of scenes such as mountains, forests, farmland, farms, and urban areas. Moreover, the shooting process occurred under diverse weather conditions, including sunlight, clouds, fog, rain, and more, reflecting substantial variations in light intensity within the images. Additionally, due to inherent randomness in shooting angles and other characteristics, the inspection target often resides at the edge of the image, further complicating the classification of these inspection images. This dataset's unique characteristics offer a valuable resource for developing and testing robust models in real-world settings, addressing the challenges posed by diverse environmental conditions and image compositions.

3.2 Dataset description

This article introduces a self-constructed inspection photo dataset, encompassing a total of 538,500 kV overhead transmission line towers. Each of these towers is composed of 13 distinct inspection parts, each captured by one or more inspection photos. In aggregate, the dataset comprises 7,516 inspection photos, after screening with an average of approximately 13.96 photos associated with each tower. Notably, the photos depicting tower inspection sites in this dataset adhere to grid inspection shooting standards and are captured from multiple angles to ensure comprehensive coverage. To facilitate organisation, inspection parts of the same class are systematically sorted into numbered folders. In contrast to other existing datasets, our inspection photo dataset exhibits several distinctive characteristics, including:

- Increased categories: The inspection photo dataset encompasses all inspection parts of tangent towers, spanning from the tower base to the towering peak. It places explicit emphasis on 500 kV transmission line towers. To our knowledge, this dataset represents a pioneering effort as it is the first to comprehensively capture every inspection part of tangent towers, and notably, it is entirely constructed using drone-view images.
- Wide range of perspectives: Our inspection photo dataset offers a diverse array of inspection photos captured from a multitude of dissimilar shooting angles. These photos encompass various shooting angles, encompass different target objects, and feature backgrounds that faithfully depict realistic settings. This diversity ensures that the dataset covers a broad spectrum of real-world scenarios and perspectives.

Figure 1 Sample images from TTower-345 (see online version for colours)

Table 1 The type of inspection parts

Label	*Inspection parts*
1	Overall appearance
2	Tower foundation
3	Tower head
4	Earthwire
5	Upper cross arm hanging point
6	Upper insulator string
7	Upper conductor hanging point
8	Middle cross arm hanging point
9	Middle insulator string
10	Middle conductor hanging point
11	Lower cross arm hanging poin
12	Lower insulator string
13	Lower conductor hanging point

In this paper, a subset of image samples from the TTower-345 dataset is showcased in Figure 1 and Table 1 provides an overview of the various types of inspection parts that are captured within the dataset.

3.3 *Training and testing dataset*

The training and test sets are partitioned in a ratio of 6.4:4.6. Specifically, the training set comprises 268 towers, encompassing 4,818 inspection images, each containing 13 distinct inspection parts. Conversely, the test set consists of 153 towers, encompassing 2,698 inspection images, also spanning 13 types of inspection parts. The configuration of the training and test sets is outlined in Table 2.

Table 2 The type of inspection parts

Category	*Training set*	*Test set*
	Total	*Total*
Overall appearance	268	151
Tower foundation	267	142
Tower head	535	291
Earthwire	536	290
Upper cross arm hanging point	267	153
Upper insulator string	268	153
Upper conductor hanging point	536	303
Middle cross arm hanging point	262	152
Middle insulator string	267	153
Middle conductor hanging point	541	304
Lower cross arm hanging point	262	152
Lower insulator string	268	151
Lower conductor hanging point	541	303

4 Automatic naming of inspection photos

4.1 *Basic knowledge of EfficientNet model*

EfficientNet, developed by Google Brain and introduced in the paper titled 'EfficientNet: rethinking model scaling for convolutional neural networks', presents an innovative approach to scaling convolutional neural networks (CNNs). It introduces the concept of compound coefficient, which efficiently scales network dimensions, including width, depth, and resolution. By leveraging this unique scaling method in conjunction with AutoML technology, EfficientNet achieves a remarkable tenfold improvement in computational efficiency compared to traditional CNN (Farag, 2017) models. EfficientNet aims to establish a standardised framework for expanding convolutional networks, achieving both high accuracy and resource conservation. Optimisation of resolution, depth, and width is carefully balanced to strike an ideal equilibrium between efficiency and accuracy (Tan and Le, 2019). Incorporating elements from other successful network architectures, EfficientNet utilises MBCConv as the primary network structure, drawing inspiration from MobileNet V2. Furthermore, it incorporates the squeeze and excitation (SE) method from squeeze and excitation networks (SENet) (Jie et al., 2018). The SE module plays a crucial role by enabling attention or focus operations within the channel dimension. This allows the model to prioritise channels with higher information content while suppressing less critical ones, contributing to improved performance and representation learning.

4.2 SCDA

SCDA is an image retrieval technique rooted in deep learning. It operates by taking fine-grained images as input and passing them through a pre-trained CNN model to extract convolutional or fully connected features. This algorithm is mainly adapted for fine-grained image retrieval tasks, where its objective is to pinpoint and emphasise the primary subjects within an image while suppressing background noise and irrelevant details. Notably, SCDA is an unsupervised method, meaning that its image features depend solely on the choice of the pre-trained model, making the selection of an appropriate pre-trained model a critical factor for SCDA effectiveness (Zhu et al., 2019).

In this paper, the EfficientNet_b0 model, a member of the EfficientNet series, serves as the designated pre-training model for extracting SCDA foreground features from inspection photos. The application of the SCDA technique serves the purpose of removing complex background information from inspection photos, allowing for a more precise focus on the salient foreground features within these images. The process involves feeding inspection photos into the EfficientNet_b0 network to perform feature extraction, resulting in primary feature maps. These feature maps then undergo an SCDA operation to generate feature vectors that emphasise the foreground of the inspection photos. Unlike some other retrieval methods, SCDA has the capability to extract deep convolutional features using only pre-trained models, enabling effective identification of the primary subjects within the images. Finally, the aggregation map is obtained by summing up all the channels within the SCDA, a process expressed by equation (1).

$$A = \sum_{n=1}^{a} S_n \tag{1}$$

Next, calculating the mean value of A can obtain $M_{i,j}$. The formula is shown in equation (2):

$$M_{i,j} = \begin{cases} 1 & \text{if } A_{i,j} > \bar{a} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

To effectively eliminate noise and background information, SCDA employs a multi-layer aggregation interface method. In the context of this paper, this method involves the aggregation of feature maps from the EfficientNet_b0 MConv6 layer with those from the top layer to derive the foreground feature vectors of the image. The precise procedure for implementing this method is detailed in formula (3), with α typically set to 0.5.

$$SCDA^{+} \leftarrow [SCDA_{MConv6}, \alpha \times SCDA_{Top}] \tag{3}$$

The structure for feature extraction in SCDA is depicted in Figure 2.

4.3 Local binary pattern

The local binary pattern (LBP) operator is a commonly employed technique for describing the texture of a local region within an image (Cheng et al., 2023). Its notable advantage lies in its invariance to both rotational and greyscale variations. The LBP operator was initially introduced by T. Ojala, M. Pietikäinen, and D. Harwood in 1994 as a means of extracting texture features from images. This feature extraction technique identifies local texture characteristics present in the image.

In the original definition of the LBP operator, an image is divided into a 3 × 3 window, with the central pixel as the threshold (Zhao, 2021a). The greyscale values of the surrounding eight pixels are then compared to this central pixel's greyscale value. If the greyscale value of a surrounding pixel exceeds that of the central pixel, the corresponding bit in the resulting eight-bit binary code is set to 1; otherwise, it is set to 0. Consequently, a sequence of eight binary digits is generated for each pixel within the 3 × 3 window, yielding an eight-bit binary code, commonly referred to as the LBP code. In total, there are 256 possible LBP codes, each reflecting texture information within the region surrounding the central pixel.

The formula for calculating LBP values is expressed as follows:

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c) \tag{4}$$

where (x_c, y_c) refers to the center pixel, i_p represents the greyscale value of the center pixel, i_c signifies the greyscale value of the surrounding pixel, and s denotes a sign function.

Since the introduction of the original LBP operator, researchers have continuously proposed various enhancements and optimisations, including variants like the circular LBP operator, LBP rotation invariant pattern, and LBP uniform pattern. The LBP operator generates an LBP 'code' for each pixel, and the original LBP feature extracted from an image, which records the greyscale values of each pixel, can be represented as an LBP 'texture map' where each pixel records its LBP value. In many applications of LBP, rather than using the LBP texture map directly, the focus shifts to using the histogram of LBP feature spectrum statistics as a feature vector for classification and recognition purposes (Garg and Dhiman, 2021). Specific steps for extracting the LBP feature vector are as follows:

1 The detection window is divided into small 16×16 pixel regions (cells).

2 For every pixel within each cell, a comparison is made between the greyscale values of the surrounding eight pixels and the greyscale value of the center pixel. If the greyscale value of a surrounding pixel exceeds that of the center pixel, the position of that pixel is assigned the value 1; otherwise, it is assigned the value 0. Consequently, this comparison among the eight neighboring pixels results in an eight-bit binary code, which represents the LBP value of the center pixel within the window.

3 Calculate the histogram for each cell based on the decimal equivalent of the LBP values. Normalise the histogram.

4 In the final step, the statistical histograms of each cell are aggregated and connected to form a feature vector. This feature vector effectively represents the LBP texture feature of the entire image.

The support vector classification (SVC) machine learning algorithm can be effectively employed for the classification of inspection photos. Leveraging LBP technology in this context offers robustness to variations in illumination, thereby mitigating interference from complex backgrounds within inspection photos. This leads to improved accuracy in the classification of inspection components. The architecture of the LBP classification branch is visually depicted in Figure 3.

4.4 *Automatically named model for inspection photos*

The proposed model enhancement entails the integration of the SCDA module before the original classifier located at the end of the EfficientNet_b0 architecture. Additionally, a residual connection is introduced to combine the output of MBConv6_4 with the top layer. Following the pooling layer, a fully connected layer with 13 dimensions is incorporated into the architecture. A comprehensive description of the modified model is presented in Figure 4.

The quality of the foreground feature vectors extracted by SCDA is heavily reliant on the quality of the feature maps produced by the pre-trained model. Although the SCDA algorithm employs the maximum-connected region selection technique to identify the model's primary region of interest, this process may inadvertently exclude other pertinent classification information that may be distributed throughout the model, potentially diminishing classification accuracy. To address potential information gaps resulting from SCDA, a machine learning model centered on LBP feature classification was incorporated in parallel with EfficientNet. This approach involves comparing the classification results obtained from both models and selecting the result with greater confidence, thereby enhancing the overall robustness of the classification model. After a comprehensive evaluation, the IELC model was chosen as the baseline model for the automatic naming of inspection photos. The complete architectural representation of IELC is presented in Figure 5.

Since auto-naming of inspection photos is a multi-classification problem, the cross-entropy loss function is used to compute the similarity between predicted and actual values. Minimising the cross-entropy loss between the network output and the corresponding labels is indicative of improved network classification. One advantage of using the cross-entropy loss function is to avoid learning rate decay experienced by the mean square error loss function due to the influence of the sigmoid function during gradient descent (Sun et al., 2020). The mathematical representation of the crossentropy function is:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{M} y_{i,c} \times \log p_{i,c} \tag{5}$$

where M represents the number of classes, N refers to the number of samples, $y_{i,c}$ is a binary function that returns a value of 1 if the true label of sample i is c (belonging to class c); otherwise, it assigns a value of 0, indicating that the sample does not belong to class c. The probability of sample i belonging to category c is denoted by $p_{i,c}$.

The LBP classification network in this context employs the classical hinge loss function (Lin et al., 2017). Hinge loss is a well-established loss function frequently used in machine learning. It is a non-convex function commonly applied during the training of support vector machine models. This loss function operates in a nonlinear manner and enhances the model's accuracy by guaranteeing that 'positive' and 'negative' samples are positioned on opposite sides of a hyperspace, ultimately contributing to the classifier's accuracy.

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq y_i} \max(0, s_j - s_{y_i} + \Delta) \tag{6}$$

The equation includes y_i, which represents the true class label, s_{y_i}, indicating the score for the true class label, and s_j, denoting the score for all other non-true class labels. These scores are associated with predicting an incorrect label. Typically, Δ is assigned a value of 1.

Figure 2 The structure of SCDA (see online version for colours)

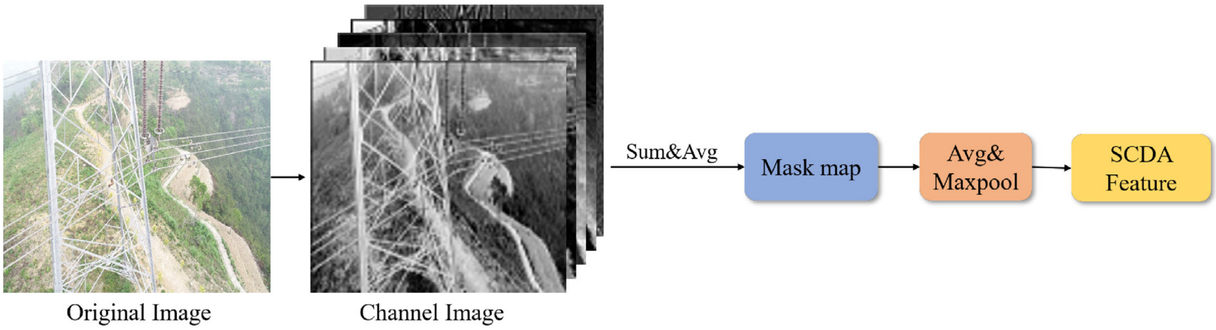


Figure 3 The structure of the LBP classification branch (see online version for colours)

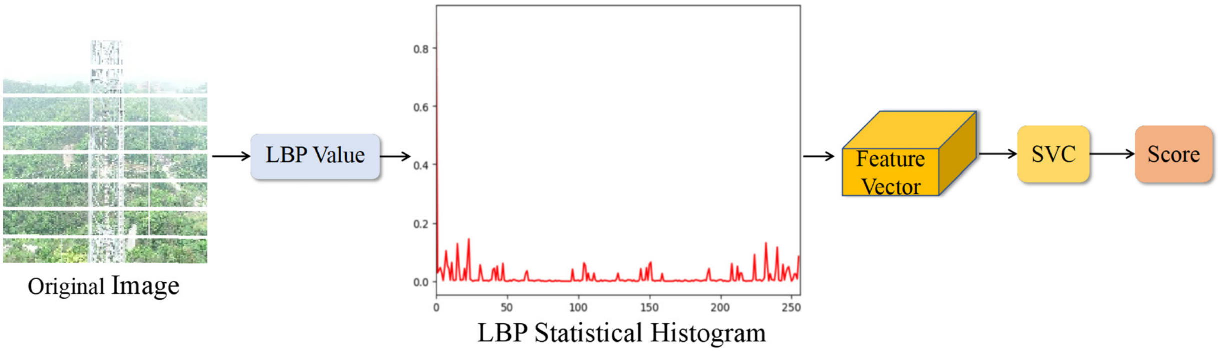


Figure 4 The structure of the improved EfficientNet framework (see online version for colours)

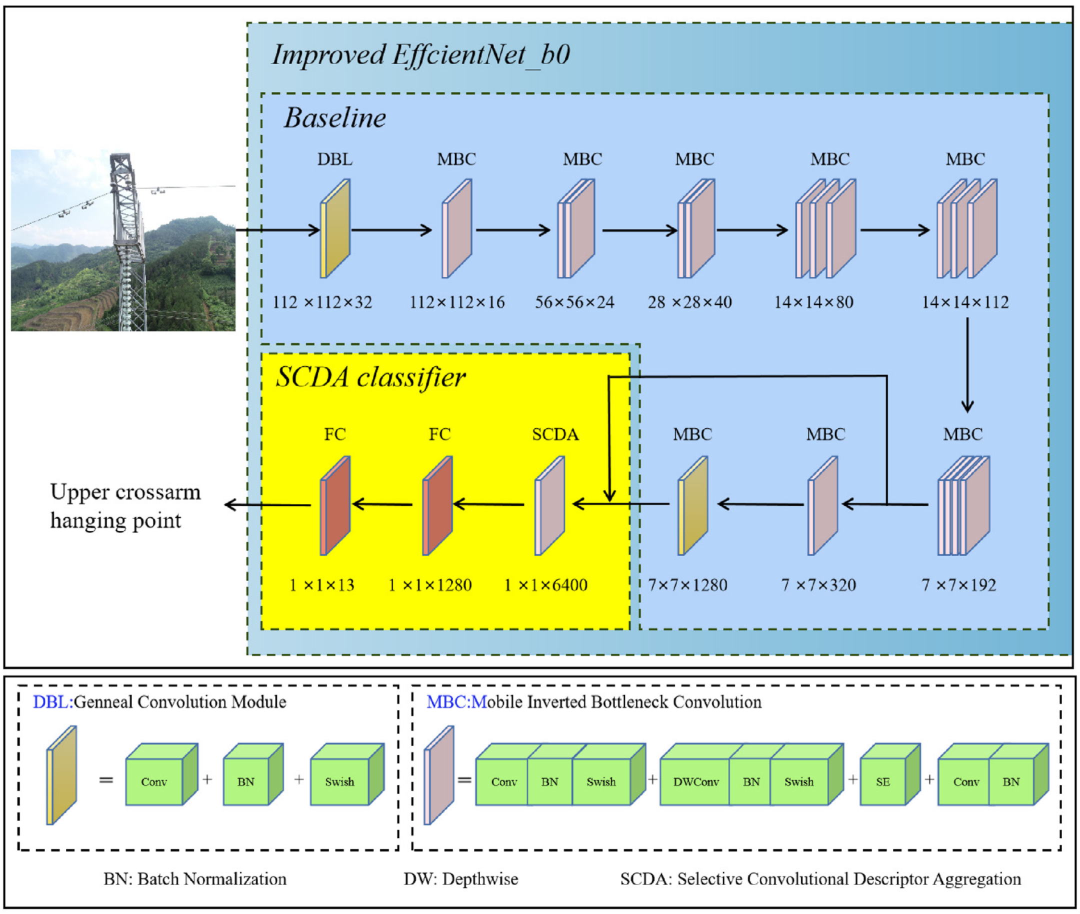
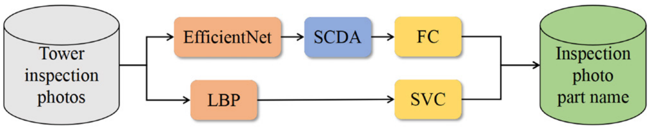


Figure 5 Automatic naming framework IELC (see online version for colours)



5 Experimental results and discussion

5.1 Experimental environment and parameter settings

In our work, we utilised the pre-trained EfficientNet_b0 network, which had previously been trained on the ImageNet dataset, as the foundational model for our study. During the training phase, we applied a consistent set of data augmentation techniques. We also performed random cropping and resizing of all input images to a size of 224 × 224 using bilinear interpolation.

Table 3 The hardware configuration and software development environment

Hardware configuation	*Version or value*	*Software development environment*	*Version*
Operating system	Windows 11-64 bit	Pycharm Community Edition	2021.3
Graphics card	NVIDIA GeForce RTX 3060 Laptop GPU	Anaconda3	5.3.1
Processor	AMD Ryzen 7 5800H with Radeon Graphics	Python	3.7.0
Operating memory	6 GB	Cuda	11.1.134

Our initial training involved fine-tuning the base EfficientNet_b0 network on the TTower-345 dataset. In this process, we employed the SCDA operation to merge the feature map from the last Conv layer of the EfficientNet_b0 network with the feature map from MConv6_4. This integration allowed us to extract feature vectors representing foreground targets in inspection images. These extracted vectors were utilised for classification through the introduction of a new 13-dimensional classification layer in our paper.

To ensure that the feature map area was correctly selected during this extraction process, we introduced an LBP feature-based classification branch. This branch was designed to balance the impact of SCDA on classification results. Specifically, the LBP feature-based classification branch utilised the LBP histogram derived from the inspection image as input and employed a trained support vector machine (SVM) for classification, with the SVM (Furukawa and Deng, 2022) model obtained from the Sklearn.SVM library. During the test phase, we relied on the EfficientNet_b0 branch to extract features. The linear classifier outputs classification scores. Simultaneously, the LBP branch constructed a histogram by extracting LBP features and produced classification scores via the SVM classifier. We then compared the two types of scores and selected the score with higher classification confidence as the final output result. This approach ensured robust classification performance.

The hardware configuration of the experimental equipment and the development environment are presented in Table 3.

Table 4 shows the training parameters for the improved EfficientNet model.

Table 4 The parameters settings for network training.

Parameters	*Value*	*Parameters*	*Value*
Input shape	224 × 224	Learning rate	0.1
Optimiser	Stochastic gradient descent	Final learning rate	0.01
Epochs	100	Lr_Lambda	Cosine
Batch size	4	Loss_Function	CrossEntropyLoss

In our training process, we implemented a training freeze method to expedite the training of the enhanced EfficientNet model. During the freezing phase, the backbone feature extraction network remained fixed, meaning its weights and parameters were not updated. Conversely, layers of the network linked to the classification head were not frozen and were iteratively adjusted during training iterations.

This approach offers several advantages. Firstly, it reduces the demand for hardware resources as fewer parameter updates are required during the freezing phase. Secondly, it mitigates the effects of random weight initialisation, which can be problematic when working with datasets that have limited sample numbers. As a result, this method significantly enhances training efficiency, accelerates model convergence, and shortens overall training duration.

5.2 Data augmentation

To mitigate the risk of underfitting during model training with the inspection photo dataset, we applied data augmentation to each training image before training the enhanced EfficientNet model. The data augmentation techniques we used included random cropping, resizing, flipping, and normalisation. Additionally, we introduced random reductions in the image size. Here are specific examples of data augmentation techniques employed, as illustrated in Figures 6(a) and 6(c).

5.3 Model evaluation

In this paper, the performance of both the improved EfficientNet and LBP models is assessed using accuracy as the primary metric. Accuracy is a crucial evaluation metric that takes into account several variables derived from the confusion matrix. The confusion matrix offers an approximate assessment of the prediction results and demonstrates the overall relationship between the original data and the predicted outcomes.

Figure 6 Examples of training image data augmentation, (a) original image (b) random cropping (c) random flipping (see online version for colours)

(a)

(b)

(c)

Table 5 Ablation experiments of the improved EfficientNet on the inspection photos dataset

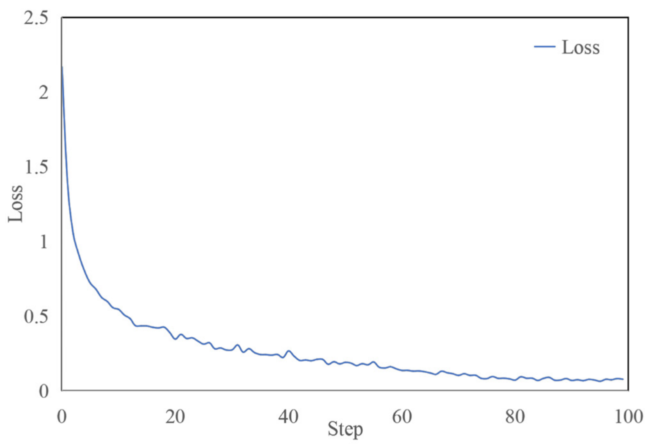
Module	*Experiment 1 result (%)*	*Experiment 2 result (%)*	*Experiment 3 result (%)*
EfficientNet_b0 backbone	√	√	√
SCDA classifier		√	√
LBP branch			√
Accuracy	95.84	97.83	*98.43*

The mathematical representation of accuracy, used for evaluating classification models, is as follows:

$$Accuracy = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \tag{7}$$

Figure 7 illustrates the training loss curve of the enhanced EfficientNet model. During the frozen training stage, the training loss experienced a swift decline, facilitating rapid model convergence. As the number of iterations increased, the training loss values steadily decreased. Eventually, the loss curve leveled off, reaching a satisfactory convergence state. This observation indicates that the model underwent effective training and achieved the desired performance.

Figure 7 The training loss curve of the improved EfficientNet model (see online version for colours)



5.4 Ablation experiment and discussion

5.4.1 The effect of improved EfficientNet model on classification results

This section delves into the outcomes of the experimental ablation tests conducted on the TTower-345 dataset, aimed at assessing the effectiveness of the SCDA module and LBP classification branch. The results of these tests are presented in Table 5.

The data in Table 5 reveal that the utilisation of the SCDA module in conjunction with the EfficientNet_b0 network led to a noteworthy 1.99% increase in classification accuracy compared to employing the EfficientNet_b0 module in isolation. These findings indicate that the SCDA module can effectively identify foreground objects of interest to the network while concurrently filtering out background image interference.

Table 6 Ablation study of different input sizes on the inspection photos dataset

EfficientNet model	*Image size*	*Accuracy (%)*	*FLOPs (M)*	*Params (M)*
EfficientNet_b0	224	*95.84*	*411.57*	4.02
EfficientNet_b0	240	94.99	492.67	4.02
EfficientNet_b0	260	92.32	613.18	4.02
EfficientNet_b0	300	91.22	778.09	4.02
EfficientNet_b0_scda	224	*97.83*	*696.81*	7.72
EfficientNet_b0_scda	240	96.21	839.74	7.72
EfficientNet_b0_scda	260	93.35	1,047.55	7.72
EfficientNet_b0_scda	300	92.43	1,325.95	7.72

Furthermore, by integrating the LBP classification branch, it is possible to recover edge detail information that might have been lost during the SCDA foreground extraction process. This resulted in an enhanced classification accuracy of 2.59% over that of the EfficientNet_b0 network used independently. It is important to note that the SCDA method, despite its design to select class-relevant pixels within the maximum connected region, can potentially lead to information loss. Therefore, the texture features and weight allocation module extracted by LBP can compensate for such losses by gathering global information that may have been lost, thereby further improving classification accuracy.

5.4.2 The effect of image resolution

In our experiments, we explored the impact of varying input image sizes on the classification accuracy of inspection images. We applied the corresponding EfficientNet model configurations listed in Table 6 for different input sizes. Based on the experimental findings, increasing the input size of the EfficientNet benchmark model to 240 results in improved classification accuracy for inspection images. However, when the input size is further increased to 300, a performance decline is observed. We hypothesise that the decreased performance with a larger input size is due to a substantial deviation from the pretrained weight size (224×224) on ImageNet. This disparity in input size could potentially impact the model's generalisation ability, leading to a degradation in performance. Furthermore, the experimental results demonstrated that despite the increase in model computation and parameter count following the incorporation of the SCDA module, for instance, at a resolution of 224×224, where the model parameter count increased by 1.7 times compared to the original, the accuracy exhibited a notable improvement of 1.99. This provides additional evidence supporting the efficacy of the SCDA module in enhancing the classification of inspection images.

5.4.3 The effect of different backbone models

To assess the performance of the model employed in this paper on the TTower-345 dataset, we trained a range of backbone network models, including VGGNet16 (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2014), DenseNet21 (Huang et al., 2016), ShuffleNet (Zhang et al., 2017), ConvNeXt (Liu et al., 2022), etc. The results of these experiments are documented in Table 7.

Table 7 Comparison of results based on different models

Model	*Result (%)*
VGGNet16	80.31
GoogLeNet (InceptionV1)	90.38
DenseNet21	94.76
ShuffleNet V2	95.10
EfficientNet_b0	*95.84*
EfficientNet V2_s	78.04
ConvNeXt_Tiny	94.61
Ours	*97.83*

The training process of different backbone network models is illustrated in Figure 10.

Based on the outcomes presented in Table 7, it is evident that the improved EfficientNet model introduced in this paper outperformed other models (He et al., 2022). Specifically, compared to the DenseNet21 model, our proposed model achieved a superior performance improvement of 3.07%. Additionally, it outperformed the ConvNeXt_tiny model by a notable margin of 3.22%, and it also surpassed the EfficientNet_b0 model by 1.99% in the task of normal inspection image classification. These impressive results affirm the suitability of the proposed EfficientNet architecture for image classification tasks on the given dataset. Furthermore, the proposed improved model demonstrated its capability to enhance the accuracy of inspection image classification, showcasing its potential as an effective solution for this task.

5.4.4 The classification results of inspection images in different scenarios

To provide further evidence of the efficacy of the proposed improved EfficientNet model, we conducted tests using inspection images captured in various background environments. The results of these tests are depicted in Figure 8. These results highlight the model's impressive capability to accurately classify the location categories of inspection images, irrespective of variations in background brightness, weather conditions, and the presence of building structures or terrain features in the background. This demonstrates the robustness and versatility of the proposed model in handling diverse environmental conditions.

Figure 8 Comparison of classification results in different background environments (see online version for colours)

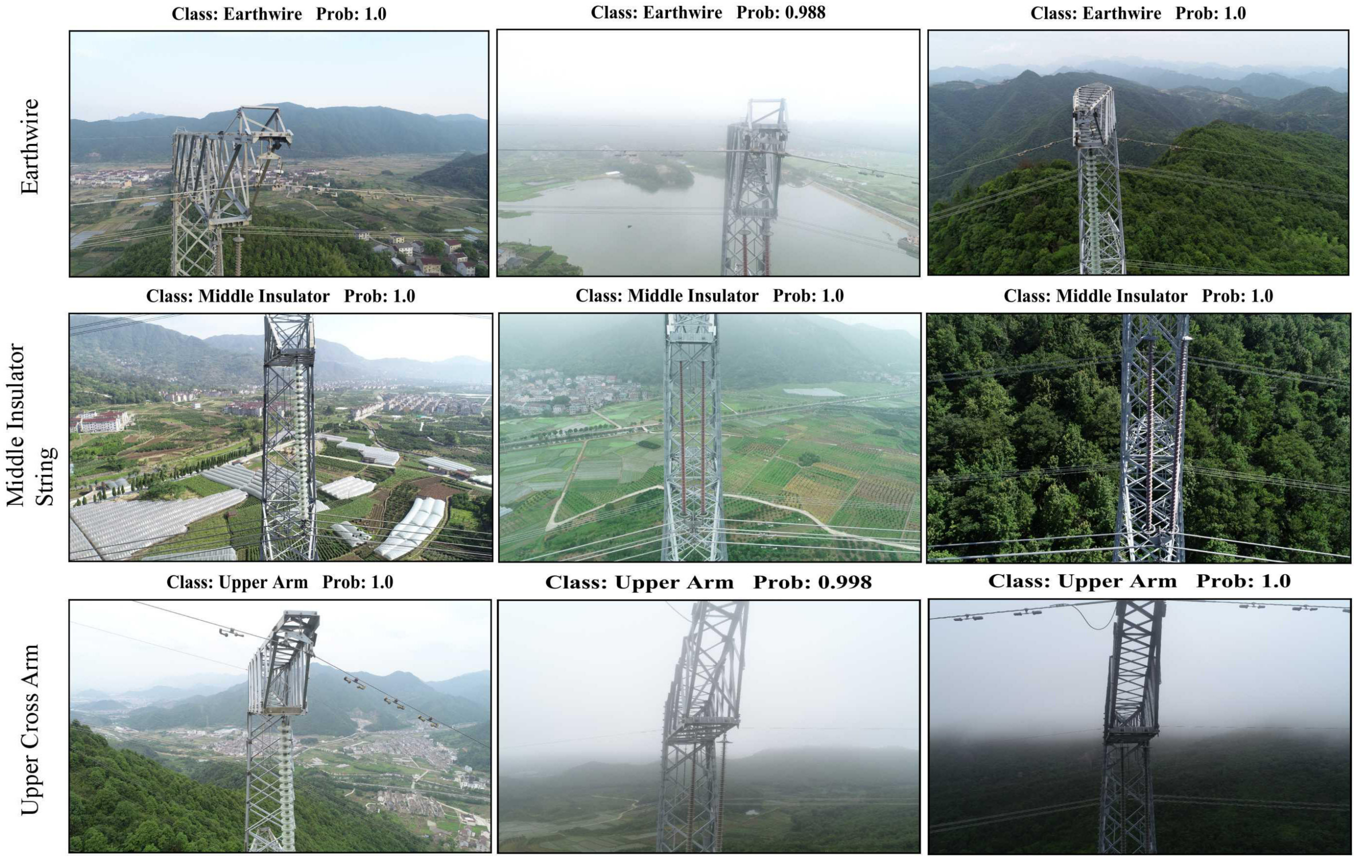


Figure 9 Classification results of different categories (see online version for colours)

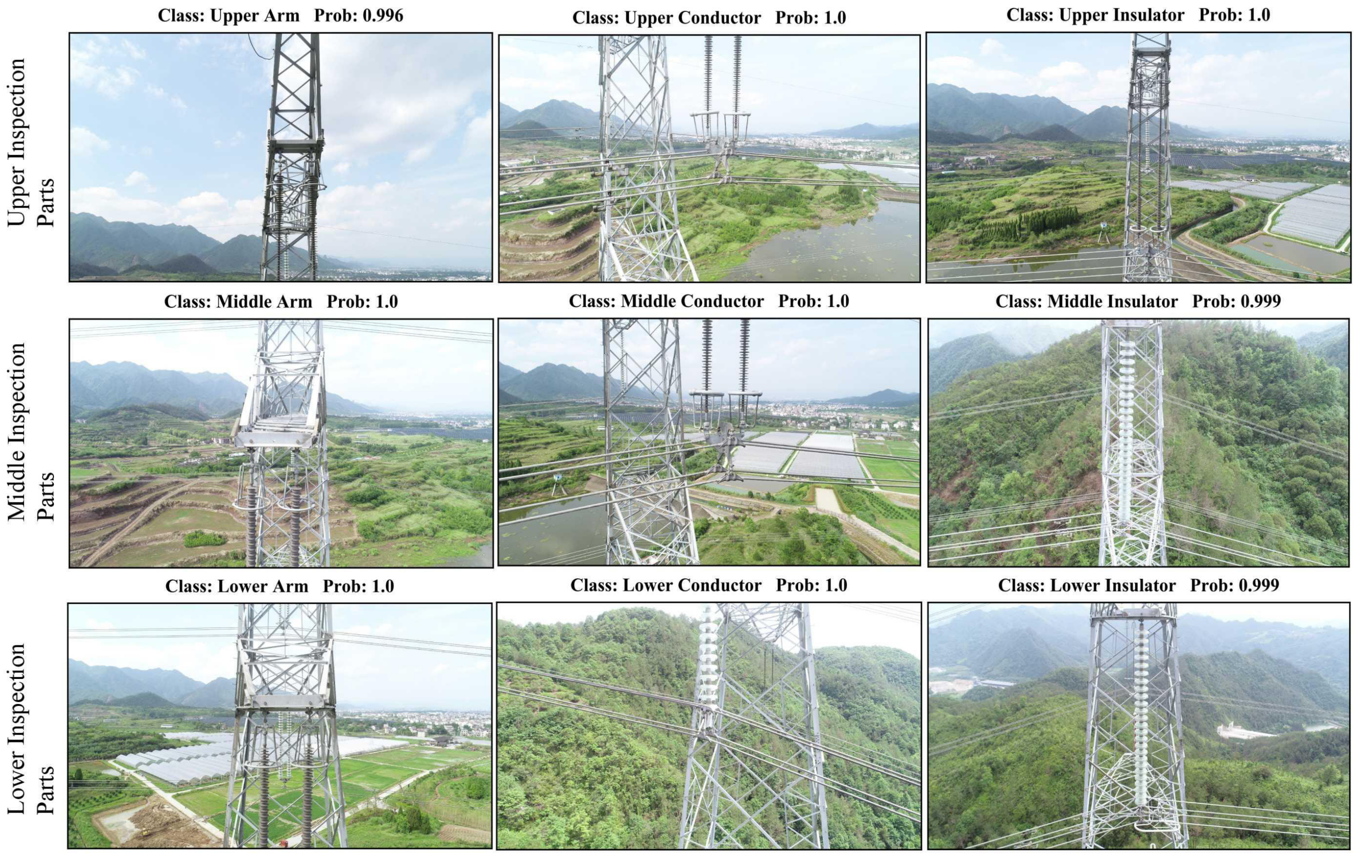
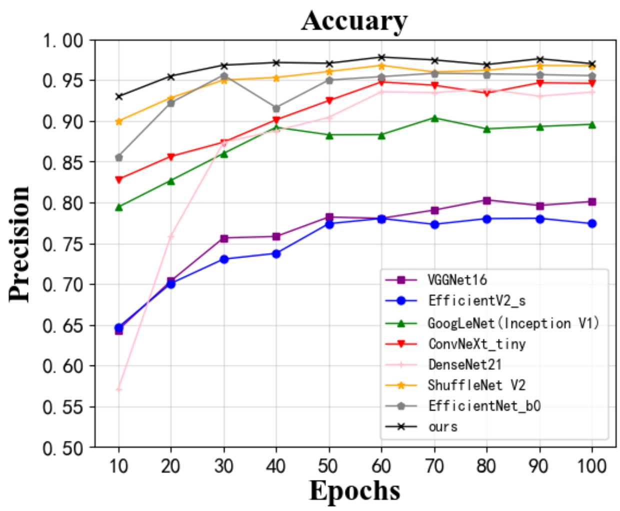


Figure 10 The training curve of different models (see online version for colours)



5.4.5 *Classification results of inspection photos with different categories*

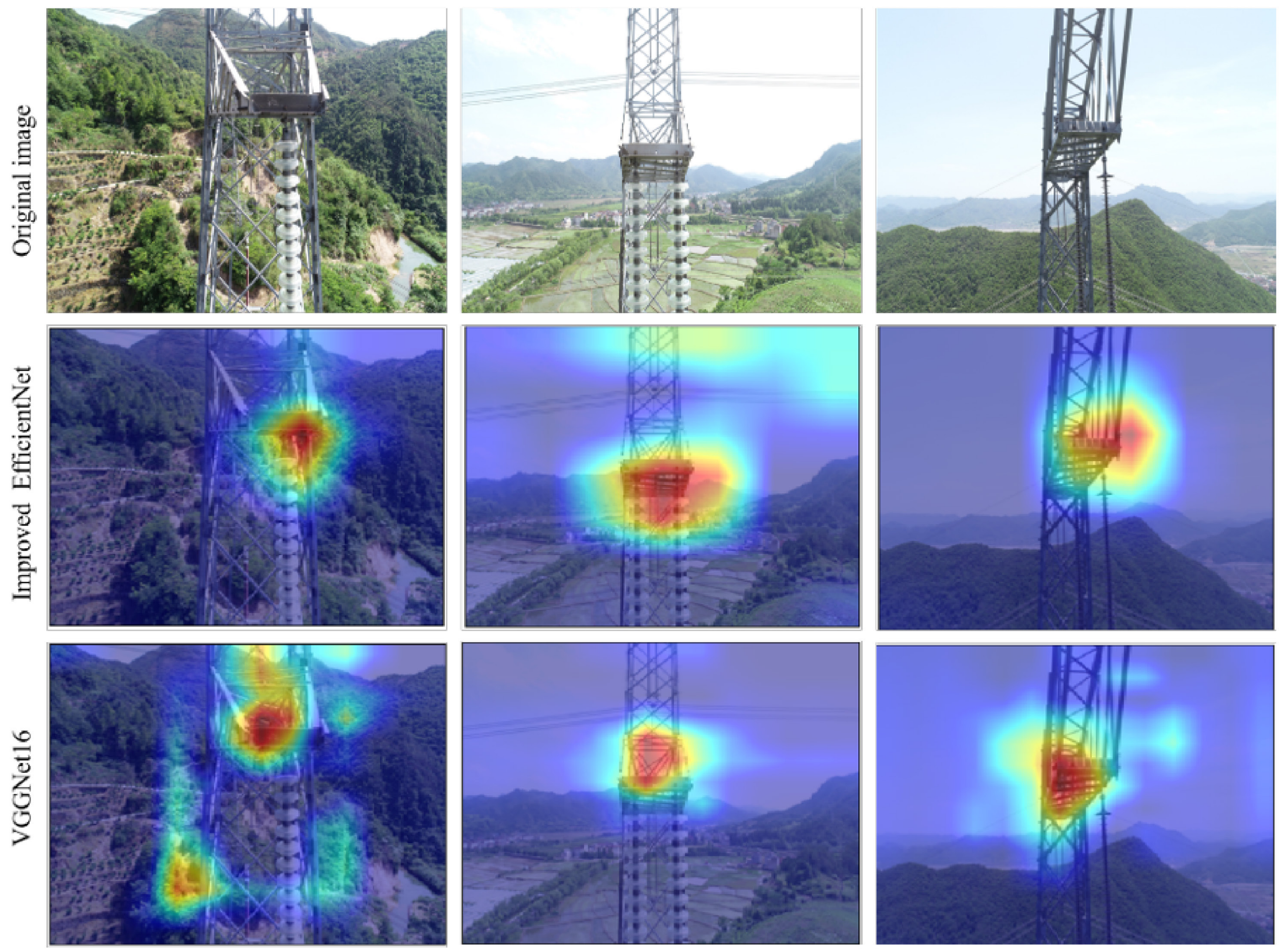
The identification of inspection features from images taken at different positions (upper, middle, and lower) on dual-circuit transmission towers can be challenging due to the small visual differences between these positions. In response to this challenge, the improved EfficientNet model proposed in this paper was applied to classify inspection images from these different sections of transmission towers.

The results demonstrate the model's high generalisability and granularity in classification, allowing it to accurately classify multiple categories of inspection images corresponding to these different positions. Specific results are visualised in Figure 9. This showcases the model's effectiveness in handling classification tasks with varying levels of granularity and complexity.

5.4.6 *Visualisation of target feature activation in classification*

The inherent opacity of deep learning networks' training and testing processes has spurred the development of visualisation tools like class activation mapping (CAM) by researchers such as Zhou et al. CAM is designed to shed light on the learning process of CNNs by generating heat maps of input images using CNNs. These heatmaps highlight regions with higher response levels, indicating areas with significant impact on classification and localisation results, and whether they are situated at the core of the target.

Figure 11 Comparison of classification results in different background environments (see online version for colours)



In the context of this paper, the reliability of the models was assessed by identifying the largest active area in the image and comparing it with the position of cable endpoints in inspection photos. Specifically, the study focused on images of cable terminal attachment points, which were input into both the EfficientNet and VGGNet16 models. Utilising the CAM tool, decision regions with higher response levels from both models were visualised, as illustrated in Figure 11.

The results revealed that the improved EfficientNet model demonstrated more accurate identification of inspection sites compared to the VGGNet16 model. Additionally, the decision region of the improved model exhibited a higher level of response. These findings underscore the improved model's superior performance in extracting foreground target features from inspection sites, further validating its effectiveness.

6 Conclusions

This paper introduces the TTower-345 dataset, a multi-category and multi-view inspection photo dataset comprising images captured by an UAV platform during tangent tower inspections. The dataset is designed for tasks such as inspection photo naming and defect detection in power grid maintenance processes. The primary focus of this study is on the image classification task for inspection photo naming.

To address this task, the paper proposes an improved EfficientNet network model that incorporates the SCDA module. This model utilises the SCDA module to extract foreground target features, building on the lightweight EfficientNet_b0 model. Several techniques are employed to enhance the model's classification accuracy, including data augmentation and increased input image resolution. Furthermore, the model's robustness is reinforced by adding an LBP classification branch, resulting in the creation of the IELC automatic naming framework for inspection photos. Experimental results, including comparisons with the backbone network and the use of the CAM visualisation tool, demonstrate that the proposed improved EfficientNet model can effectively learn competitive features from inspection photos. The IELC framework, in conjunction with the integrated LBP classification model, achieves a high overall classification accuracy, currently reaching an impressive rate of 98.43%. This level of accuracy satisfies the naming requirements for inspection photos of tangent towers in the power grid sector. However, it is noted that the use of multiple parallel models can lead to longer naming times. In future work, there should be a consideration for developing more lightweight naming structures to enhance practical application efficiency. Researchers can leverage the TTower-345 dataset to create faster and more accurate models for automatic photo inspection naming and defect detection, thereby contributing to the development of intelligent information management platforms for the power grid to address real-world needs.

Acknowledgements

This work was funded by Jinhua Bada Group Co., Ltd. Science and Technology Project No. BD2022JH-KXXM007.

References

Cheng, Q., Li, X., Zhu, B., Shi, Y. and Xie, B. (2023) 'Drone detection method based on MobileViT and CA-PANET', *Electronics*, Vol. 12, No. 1, p.223.

Du, Q., Dong, W., Su, W. and Wang, Q. (2022) 'UAV inspection technology and application of transmission line', in *2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*, IEEE, pp.594–597.

Farag, W. (2017) 'Safe-driving cloning by deep learning for autonomous cars', *International Journal of Advanced Mechatronic Systems*, Vol. 7, No. 6, pp.390–397.

Furukawa, Y. and Deng, M. (2022) 'SVM-based fault detection for double layered tank system by considering changefinder's characteristics', *International Journal of Advanced Mechatronic Systems*, Vol. 9, No. 4, pp.185–192.

Gao, X., Yang, Q. and Zhang, J. (2022) 'Multi-objective optimisation for operator-based robust nonlinear control design for wireless power transfer systems', *International Journal of Advanced Mechatronic Systems*, Vol. 9, No. 4, pp.203–210.

Garg, M. and Dhiman, G. (2021) 'A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants', *Neural Computing and Applications*, Vol. 33, No. 4, pp.1311–1328.

Guowei, X., Yifeng, P. and Qinglei, J. (2020) 'Research and application of photograph collection method for transmission and distribution lines based on POS information and spatial position relationship of towers', *Electric Engineering*, in Chinese, Vol. 40, No. 8, pp.97–100.

He, B., Li, L., Ren, S. and Qian, D. (2022) 'Scratch detection system of the inner surface of super long gas cylinder based on VGG-16 neural networks', *International Journal of Advanced Mechatronic Systems*, Vol. 9, No. 4, pp.193–202.

Huang, G., Liu, Z., Laurens, V. and Weinberger, K.Q. (2016) 'Densely connected convolutional networks', *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society*, pp.2261–2269.

Jie, H., Li, S. and Gang, S. (2018) 'Squeeze-and-excitation networks', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kehui, Z., Zhiwei, L., Yiyao, X. et al. (2019) 'Construction of infrared image classification model for power equipments based on improved CNN', *Infrared Technology*, Vol. 41, No. 11, pp.1033–1038.

Lin, T-Y., Goyal, P., Girshick, R., He, K. and Dollár, P. (2017) 'Focal loss for dense object detection', in *Proceedings of the IEEE International Conference on Computer Vision*, pp.2980–2988.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T. and Xie, S. (2022) *A ConvNet for the 2020s*, arXiv e-prints.

Mn, A., Scaria, C.T., Thomas, S.R., Menon, S.S., Joy, S. and Chacko, G. (2022) 'Development of a robotic system for power transmission line inspection', in *2022 IEEE 19th India Council International Conference (INDICON)*, IEEE, pp.1–6.

Panigrahy, S. and Karmakar, S. (2022) 'Research on transmission line insulator defects detection using YOLOv7', in *2022 IEEE 6th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON)*, IEEE, pp.384–388.

Simonyan, K. and Zisserman, A. (2014) *Very Deep Convolutional Networks for Large-Scale Image Recognition*, arXiv preprint arXiv:1409.1556.

Su, Y. (2020) 'Development and application of a standardized input tool for transmission line images', *Electric Engineering*, in Chinese, Vol. 40, No. 6, pp.82–83.

Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z. and Wei, Y. (2020) 'Circle loss: a unified perspective of pair similarity optimization', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.6398–6407.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P. and Rabinovich, A. (2014) 'Going deeper with convolutions', in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, pp.1–9.

Tan, M. and Le, Q.V. (2019) 'EfficientNet: rethinking model scaling for convolutional neural networks', in *International Conference on Machine Learning, PMLR*, pp.6105–6114.

Tang, R., Zhang, M., Xu, H., Li, W. and Li, W. (2021) 'Insulator recognition in transmission line inspection images based on deep learning', *Power Grid and Clean Energy*, Vol. 37, No. 4, pp.41–46.

Tao, X., Zhang, D., Wang, Z., Liu, X., Zhang, H. and Xu, D. (2018) 'Detection of power line insulator defects using aerial images analyzed with convolutional neural networks', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 50, No. 4, pp.1486–1498.

Tianjiao, P., Ji, Q., Xiao, H., Guobin, Z. and Xinying, W. (2020) 'Research and application of artificial intelligence in operation and maintenance for power equipment', *High Voltage Engineering*, Vol. 46, No. 2, pp.369–383.

Wang, J., Li, Y. and Chen, W. (2023) 'UAV aerial image generation of crucial components of high-voltage transmission lines based on multi-level generative adversarial network', *Remote Sensing*, Vol. 15, No. 5, p.1412.

Xia, H., Yang, B., Li, Y. and Wang, B. (2022) 'An improved centernet model for insulator defect detection using aerial imagery', *Sensors*, Vol. 22, No. 8, p.2850.

Xiren, M., Zhiying, L. and Qichen, Y. (2020) 'Review on UAV intelligent technology for transmission line inspection', *Journal of Fuzhou University (Natural Science Edition)*, Vol. 48, No. 2, pp.198–209.

Zhang, X., Zhou, X., Lin, M. and Sun, J. (2017) 'ShuffleNet: an extremely efficient convolutional neural network for mobile devices', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.6848–6856.

Zhang, Z-D., Zhang, B., Lan, Z-C., Liu, H-C., Li, D-Y., Pei, L. and Yu, W-X. (2022) 'FINet: an insulator dataset and detection benchmark based on synthetic fog and improved YOLOv5', *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, pp.1–8.

Zhao, Q. (2021a) 'Research on the application of local binary patterns based on color distance in image classification', *Multimedia Tools and Applications*, Vol. 80, No. 18, pp.27279–27298.

Zhao, Z. (2021b) 'Insulatorgan: a transmission line insulator detection model using multi-granularity conditional generative adversarial nets for UAV inspection', *Remote Sensing*, Vol. 13, No. 19, p.3971.

Zhu, X., Pang, J., Yang, C., Shi, J. and Lin, D. (2019) 'Adapting object detectors via selective cross-domain alignment', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.687–696.