IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A Sampling based Generic Methodology for Object Detection using Feature Scale Learning

[1]UPPU PALAKONDA RAJU,[2]VADLAMUDI R V S RITWICK,[3]B.KARTHIK KUMAR REDDY, M CHITRA[4]

[1]STUDENT,[2]STUDENT,[3]STUDENT,4ASSISTANT PROFESSOR
[1]COMPUTER SCIENCE AND ENGINEERING,
[1]SRMIST, CHENNAI,INDIA

Abstract: Due to their limited pixel count, small objects present significant challenges in feature extraction, which is essential for accurate detection. They are less prevalent in datasets and often overwhelmed by many background instances, potentially biasing the training process. Additionally, as the stride of feature maps increases in deeper network layers, the spatial resolution decreases, causing small objects to be missed and critical details to be lost. Moreover, small objects often depend on contextual clues for accurate identification, which can be difficult to effectively capture and utilize.

*I.*INTRODUCTION

Object detection, which merges image classification with precise localization, is crucial for understanding images comprehensively., enabling the extraction of semantic and intricate features. This field is subdivided into various categories such as face, pedestrian, and skeleton detection, providing rich semantic information for images and videos. Despite recent advancements facilitated by deep learning, detecting small objects—defined in the COCO dataset as those occupying less than 32x32 pixels—remains challenging. Small objects provide fewer semantic features due to their limited area, are less numerous which may skew the model's focus towards larger objects and pose difficulties in anchor matching due to low Intersection over Union (IoU) values with the set anchors, which can cause the network to misclassify these anchors as negative samples. Object detection is not only one of the core problems in computer vision but also influences other research areas within the field. Its advancements are pivotal and find applications in autonomous driving, video surveillance, and more, enhancing everyday life and operations across various sectors including defence.

However, the effectiveness of object detection can vary significantly across different scenarios, and while improvements to its basic components can enhance overall performance, their impact on retrained models can be complex. With object detection being fundamental to tasks like image segmentation and object tracking, the development of both one-stage and two-stage detectors has been significant. Algorithms such as R-CNN, Faster R-CNN, and Mask R-CNN have seen improvements in accuracy, though their architecture still poses limits on detection speed. YOLO, a notable object detection algorithm using CNNs like Darknet-53 and CSPDarknet53 for feature extraction in its various versions, showcases rapid detection capabilities, though it struggles with translation variance, locality sensitivity, and lacks a global image understanding. CNN-based models typically use pooling layers to reduce dimensionality and computational costs, yet this can lead to significant information loss, particularly affecting the precise localization of critical features.

*II.*RELATED WORK.

Title: "SIMPL: Generating Synthetic Overhead Imagery to Address Custom Zero-Shot and Few-Shot Detection Problems". Bohao Huang, Yang Xu, Xiong LLuo,Kyle BBradbury,and Jordan M. Malof, Published in 2022. Recently Deep Neural Networks have achieved remarkable progress in object detection using overhead images, such as satellite photos. However, acquiring and annotating satellite imagery for training remains prohibitively expensive. To tackle the challenges of data acquisition in satellite imagery analysis, we introduce a technique known as Synthetic object Implantation. This method efficiently creates a large volume of synthetic overhead training data for specific target objects. In this study, we evaluate the effectiveness of using synthetic imagery generated by Synthetic object IMPLantation to train Deep Neural Networks in zero-shot scenarios—where no real imagery is available—and few-shot scenarios—where real-world imagery is scarce. Our findings indicate that synthetic data from Synthetic object IMPLantation provides a viable and cost-effective alternative, offering a solution to the substantial challenges associated with data acquisition in the field of satellite imagery analysis.

Title: "Few-Shot Object Detection With Self-Adaptive Global Similarity and Two-Way Foreground Stimulator in Remote Sensing Images ". Yuchen Zhang , Bo Zhang and Bin Wang, Published in 2022. Few-shot object detection (FSOD) aims to localize and recognize potential objects of interest only by using a few annotated data, and it is beneficial for remote sensing images (RSIs) based applications, such as urban monitoring. Previous RSIs-based FSOD works often try to convert the support images from class-agnostic features to class-specific vectors, and then perform feature attention operations on query image features to be tested. However, such methods still face two critical challenges: 1) They ignore the spatial similarity of support-query features, which is indispensable for RSIs detection; 2) They perform the feature attention operation in a unidirectional manner, which means that the learned support-query relations are asymmetric.

Title: "CLFM: Few-Shot Object Detection via Low-Resource Contrastive Learning and Fisher Matrix Updating for Overcoming Catastrophic Forgetting". The Authors are Meng Wang , Qiang Wang and Haipeng Liu.Published in 202. Few-shot object detection (FSOD) aims to efficiently detect novel instances by model transferring using a few novel-class samples after the base-class samples are pre-trained. However, catastrophic forgetting occurs when FSOD transfers to the novel-classes, making the transferred model unable to accurately detect base and novel class instances simultaneously. Thus, this paper attempts to extend the general elastic weight curing (EWC) to the field of few-shot transfer detection. An online soft constraint is applied by evaluating the Fisher information matrix of inner-batch samples based on mean squared error (MSE) metric to constrain or encourage model parameters transferring. Also, a momentum update based inter-batch storage mechanism is proposed to alleviate the memory strain caused by the previously applied contrastive learning modules when performing numerous contrastive encoding procedures.

Title: "Few-Shot PCB Surface Defect Detection Based on Feature Enhancement and Multi-Scale Fusion”.The authors Haodong Wang , Jun Xie , Xinying Xu and Zihao Zheng He.Published in 2022. In printed circuit board (PCB) defect detection, it is difficult to collect defect samples, and the detection effect is poor due to the lack of data. Based on the few-shot learning method, a few-shot PCB defect detection model is proposed. This model introduces feature enhancement module and multi-scale fusion module. The feature enhancement module based on the improved convolution block attention module (CBAM) can highlight the key areas of the received feature maps and suppress the interference of useless information. Aiming at the small size of PCB defects, a multi-scale feature fusion strategy is proposed. It can extract multi-scale feature maps of PCB and fuse them into a high-quality feature map containing different scale information, which can improve the detection precision of the model for small object defects.

Title: " Siamese Neural Network Based Few-Shot Learning for Anomaly Detection in Industrial CyberPhysical Systems ". Xiaokang Zhou , Wei Liang , Shohei Shimizu , Jianhua Ma and Qun Jin. It is Published in 2021. With the increasing population of Industry 4.0, both AI and smart techniques have been applied and become hotly discussed topics in industrial cyber-physical systems (CPS). Intelligent anomaly detection for identifying cyber-physical attacks to guarantee the work efficiency and safety is still a challenging issue, especially when dealing with few labeled data for cyberphysical security protection. In this article, we propose a few-shot learning model with Siamese convolutional neural network (FSL-SCNN), to alleviate the over-fitting issue and enhance the accuracy for intelligent anomaly detection in industrial CPS. A Siamese CNN encoding network is constructed to measure distances of input samples based on their optimized feature representations. A robust cost function design including three specific losses is then proposed to enhance the efficiency of training process. An intelligent anomaly detection algorithm is developed finally.

Title: " Few-Shot Ship Classification in Optical Remote Sensing Images Using Nearest Neighbor Prototype Representation". The Authors are Jiawei Shi , Zhiguo Jiang and Haopeng Zhang. It is Published in 2021. With advances in ship detection using optical remote sensing images, obtaining accurate detection results and images of ships has become more straightforward. The use of convolutional neural networks (CNNs) is a common approach for classifying different types of ships by training a classifier with a large collection of ship images. However, the performance of CNNs can diminish when only a small number of training samples are available. To address this issue, we propose a metric-based few-shot learning method that generates novel concept representations for ship classes using a nearest neighbor prototype approach. This method differs from traditional few-shot methods, which typically rely on image-to-image measures. Instead, we employ an image-to-feature measure, enhancing the model's ability to generalize from limited data by focusing on feature similarities rather than direct image comparisons

Title: " Few-Shot Object Detection With Self-Adaptive Attention Network for Remote Sensing Images”.The Authors are Zixuan Xiao , Jiahao Qi , Wei Xue and Ping Zhong. It is Published in 2021In the field of remote sensing, object detection has seen extensive applications and typically requires a large amount of labeled data. However, scenarios often arise where only limited data are available. To address this challenge, we have developed a few-shot object detector specifically designed to identify novel objects using only a few examples. Uniquely, our detector focuses on the relationships at the object level rather than across the entire image. This is facilitated by a self-adaptive attention network (SAAN) that enhances object-level relations through a relation gate recurrent unit. The SAAN dynamically focuses on object features based on these relations, steering clear of scenarios where additional attention might be redundant or harmful. The attention-enhanced features are then used to produce detection results, simplifying the detection process in few-shot conditions. Our experiments validate the effectiveness of our proposed method in these scenarios, demonstrating its practical utility in remote sensing applications where data limitations are common.

Title: " Insulator Anomaly Detection Method Based on Few-Shot Learning ". Zhaoyang Wang , Qiang Gao , Dong Li , Junjie Liu , Hongwei Wang , Xiao Yu and Yipin Wang. Due to the advantages of safety and economy, it has become a trend to use unmanned aerial vehicles (UAVs) instead of humans to inspect high-voltage transmission lines. Considering the manual inspection process and the few-shot learning, a two-stage method for insulator anomaly detection is proposed. In the first stage, a positioning-restoration-cropping method is discussed for insulator string detection and processing. In the second stage, an insulator anomaly detection model called a multi-scale feature reweighting (MFR) network is built. With the help of few-shot object detection, the detection of five kinds of anomaly insulator caps, such as falling off, breakage and ablation is realized. The mean average precision (mAP) of the proposed method is 88.76%.This paper presented a method to detect the anomaly insulator caps based on few-shot object detection. The proposed two- stage architecture can locate insulators in aerial images and detect ve common insulator anomalies through the support set. The method retained the pixel information through image cropping, which enabled the model to detect more complex and subtle defects than the insulator caps falling off. Also, the multi-scale feature weighting network which is structured based on few-shot object detection could make full use of the feature information in the support set and solve the problem of long-tail distribution.

Title: " Few-Shot Object Detection via Sample Processing". Honghui Xu , Xinqing Wang , Faming Shao , Baoguo Duan and Peng Zhang. It is Published in year 2021. Few-shot object detection (FSOD) eliminates the dependence on tremendous instances with manual annotations in conventional object detection. We deem that the scarcity of positive samples is the main reason that restricts the performance of FSOD detectors. In this paper, a novel FSOD model via sample processing, namely, FSSP, is proposed to detect objects accurately with only a few annotated samples, which is based on the structural design of the Siamese network and uses YOLOv3-SPP as the baseline. Central to FSSP are our designed self-attention (SAM) and positive-sample augmentation modules. The former attempts to better extract the representative features of hard samples, latter expands the number and enriches the scale distribution .

Title: " Meta-SSD: Towards Fast Adaptation for Few-Shot Object Detection With Meta-Learning". The Authors are Kun Fu , Tengfei Zhang , Yue Zhang , Menglong Yan , Zhonghan Chang , Zhengyuan Zhang and Xian Sun. It is Published in Year 2020. State-of-the-art object detection frameworks typically require extensive training on large datasets, which can introduce challenges such as overfitting, decreased effectiveness when limited samples are available, and prolonged training durations. To address these issues, this paper presents a generalized Few-Shot Detection framework that utilizes meta-learning. This framework includes a meta-learner and an object detector that work together to rapidly acquire general knowledge and develop swift adaptation strategies for multiple tasks. The meta-learner instructs the object detector on how to efficiently learn from a few examples within a single update step. Although this framework could theoretically integrate any supervised learning detection model, this particular study employs the Single-Shot MultiBox Detector, resulting in the naming of the framework as Meta-Single-Shot MultiBox Detector. Furthermore, a novel benchmark derived from the Pascal Visual Object Classes dataset was established to train and evaluate the meta-learning-based Few-Shot Detection approach. Experimental findings indicate that Meta-Single-Shot MultiBox Detector achieves promising results in Few-Shot Detection scenarios. In addition, a thorough analysis of Meta-Single-Shot MultiBox Detector's features establishes a solid baseline and offers valuable insights for further research into meta-learning applications in Few-Shot Detection.

*III.*RESEARCH METHODOLOGY

The methodology section outline the plan and method that how the study is conducted. This includes Universe of the study, sample of the study,Data and Sources of Data, study’s variables and analytical framework. The detailsare as follows;

3.1MODULE 1

The Feature Pyramid Network (FPN) is a prominent architectural model in object detection that produces hierarchical Feature Pyramid Network excels in creating pyramidal feature representations crucial for detecting objects at various scales. It utilizes a backbone model, generally intended for image classification, to construct a feature pyramid. This construction is achieved by strategically merging two adjacent layers within the model's feature hierarchy. The merger is enabled through top-down and lateral connections. The process includes upsampling high-level features, which are rich in semantic information but low in resolution, and combining them with lower-level features that offer higher resolution. This integration produces feature representations that are both high-resolution and semantically robust, significantly enhancing the model’s ability to detect objects across different scales.

3.2 MODULE 2

The prediction head \(h \) in object detection architectures such as FCOS and RetinaNet plays a vital role by mapping each level of the feature pyramid \(P \) to its corresponding output \(y \). Traditionally, this process involves a series of four 3x3 convolutions. To further explore and optimize the head's capabilities, we have introduced an extended sequential search space for designing the prediction head. This new configuration includes six basic operations, enhancing the head's flexibility and adaptability. In our design, we make two significant modifications to the standard approach: Firstly, we incorporate standard convolution modules, such as 1x1 and 3x3 convolutions, into our sampling pool to allow a broader comparison and facilitate the design of more effective convolutional structures. Secondly, we replace all Batch Normalization (BN) layers with Group Normalization (GN) following the FCOS model's practices. This adjustment is made to accommodate the weight-sharing across different levels of the pyramid, where BN is less effective due to variations in mini-batch sizes. The final design outputs from the sixth layer of the head, providing a robust and semantically rich feature representation tailored for precise object detection across various scenarios.

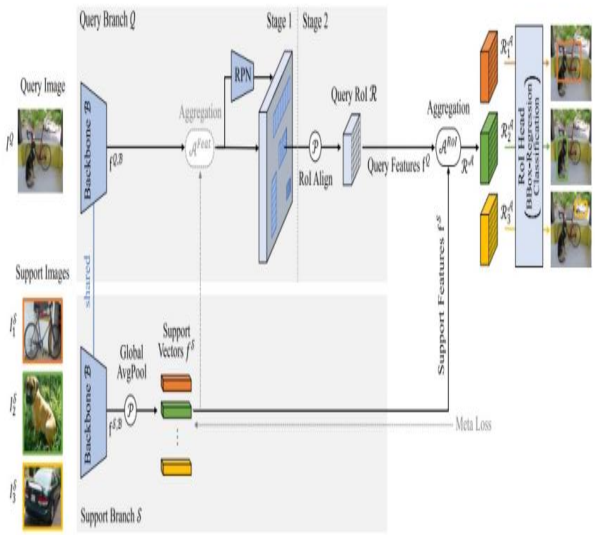
3.3 MODULE 3

To enhance the flexibility of prediction heads and delve deeper into the effects of weight sharing, we introduce an index \(i \), which determines where weight sharing begins within the prediction head. In this structure, each layer before stage \(i \) uses a unique set of weights for every output level of the Feature Pyramid Network (FPN), ensuring that each layer can tailor its processing to the specific needs of each FPN level. From stage \(i \) onwards, the prediction head shifts to using shared global weights. This division of the head into two parts—the independent segment serving as an extended branch of the FPN and the shared segment operating with an adaptive length—allows us to balance the computational load effectively. The independent layers focus on extracting level-specific features that are crucial for detecting objects across various scales and contexts, while the shared layers utilize commonalities across all levels, enhancing overall model efficiency and coherence. This strategic partitioning not only boosts the adaptability and performance of the object detection system but also provides valuable insights into the benefits and mechanics of weight sharing in complex neural network architectures.

3.4 MODULE 4

To increase the flexibility and understand the impact of weight sharing within prediction heads, we introduce an index \(i\) that specifies the starting point for weight sharing in the prediction head structure. Before this stage, each layer operates with its own set of weights tailored to each specific output level of the Feature Pyramid Network (FPN), facilitating precise, level-specific feature extraction. Starting from stage \(i\), the prediction head shifts to using a shared set of global weights across all layers that follow. This bifurcation serves a dual purpose: the independent section acts as an extended branch of the FPN, dedicated to extracting unique features at each level, while the shared section streamlines the process by utilizing common features across different levels. This strategic division not only improves the efficiency and coherence of the model but also significantly enhances its adaptability and performance in detecting objects across various contexts and scales, offering a deeper insight into the dynamics of weight sharing in neural networks.

IV.SYSTEM ARCHITECTURE



V.DATASET

COCO (Common Objects in Context) is a comprehensive dataset used for object detection, segmentation, and captioning. It features a wide array of data including object segmentation and recognition in various contexts, as well as superpixel stuff segmentation. The dataset comprises over 330,000 images, of which more than 200,000 are labeled. It contains approximately 1.5 million object instances across 80 object categories and 91 stuff categories. Each image in the dataset is accompanied by 5 descriptive captions, enhancing its utility for tasks that require contextual understanding. Additionally, the dataset includes annotations for 250,000 people, marked with keypoints to assist in detailed human pose estimation. This rich set of features makes COCO one of the most versatile and widely used datasets in computer vision research and applications

VI.PROPOSED ALGORITHM

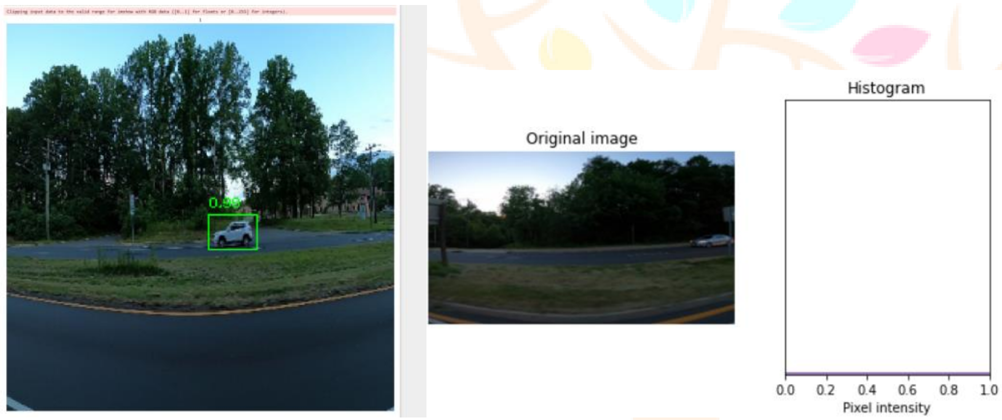
RetinaNet is a sophisticated one-stage object detection model that has gained recognition for effectively addressing one of the most persistent challenges in the field: class imbalance during training. Developed with the core aim of improving detection performance, particularly for hard-to-detect objects, RetinaNet incorporates a unique focal loss function. This function adjusts the standard cross-entropy loss by introducing a modulating term specifically designed to shift the focus towards hard negative examples. Such examples are typically overlooked in traditional detection systems due to their abundance compared to positive examples, leading to less effective learning.The architecture of RetinaNet consists of several integral components. Central to its design is the backbone network, which is generally a pre-trained convolutional network such as ResNet. This backbone is tasked with computing a rich convolutional feature map from the entire input image. Overlaying this backbone is a Feature Pyramid Network, which enhances the model's ability to detect objects at multiple scales. The FPN achieves this by creating a layered pyramid that integrates high-resolution, semantically weaker features from lower layers with the semantically stronger, lower-resolution features from higher layers through a combination of top-down and lateral connections.Building upon this robust feature base, RetinaNet employs two dedicated subnetworks: a classification subnet and a box regression subnet. The classification subnet is tasked with identifying whether objects exist at each spatial location on the feature map. It operates by predicting the probability of object presence, which guides where the model should focus its detection efforts. The box regression subnet, on the other hand, predicts the offsets needed for each anchor box to match the ground-truth object boxes as closely as possible. This dual-subnet setup allows RetinaNet to perform dense detection, which is key to its high performance.

A specific application where RetinaNet shows significant utility is in the detection of lesions in medical imaging. In such cases, the model is trained not only to distinguish between lesions and the background within the bounding boxes but also to handle the inherent ambiguities related to the boundaries of the lesions. This training involves minimizing penalties for incorrect detections while allowing some flexibility for positional adjustments of the bounding boxes, catering to the vague nature of lesion boundaries.The focal loss function plays a crucial role in RetinaNet's efficiency. By reducing the relative loss for easy negatives and increasing the importance of correcting misclassifications of difficult negatives, focal loss ensures that the training process does not become biased towards the majority class of easy negatives. This adjustment leads to a more balanced model that is better at detecting less obvious, harder-to-detect objects.Furthermore, the design of RetinaNet, with its robust backbone, sophisticated FPN, and focused loss function, makes it highly adaptable to various other applications beyond medical imaging, such as surveillance and autonomous vehicle navigation. Its ability to perform effectively across different environments and object scales is a testament to its architectural strengths.Overall, RetinaNet stands out in the landscape of object detection models due to its innovative approach to solving class imbalance with focal loss, its effective multi-scale detection capabilities enabled by the FPN, and its precise and reliable performance in detecting both small and large objects across various application domains. This makes it a valuable tool for both academic research and practical applications in computer vision.

*VII.*CONCLUSION

Object detection is crucial in surveillance systems and represents a significant application in computer vision. Traditionally, models based on Convolutional Neural Networks (CNNs) have been developed for this purpose, but they often struggle with overfitting and are inefficient at detecting smaller objects. To overcome these limitationsThis paper introduces a new target detection algorithm that improves the feature pyramid network, which we call Attention FPN.. This enhanced model integrates an improved receptive field module to better capture both global and local context information, thereby improving scene comprehension. It also includes a channel attention module in its lateral connections to highlight key features significantly contributing to object detection. Additionally, deconvolution is employed instead of traditional nearest neighbor interpolation to minimize information loss during up-sampling. Lastly, a spatial attention mechanism is applied to effectively integrate various characteristic layers, prioritizing critical information within the feature maps. These modifications collectively enhance the detection accuracy and robustness of the model, making it particularly effective in surveillance scenarios with small or complex objects.

*VIII.*RESULTS



VII.FUTURE WORK

Despite the promising aspects of our proposed algorithm, it exhibits certain limitations such as a 10% increase in the number of parameters compared to the original algorithm, and persistent background noise that hampers detection efficiency. To address these issues, our future work will focus on reducing the network's parameter count through methods like pruning and quantization, and enhancing noise robustness using advanced signal processing techniques. Additionally, we plan to integrate our algorithm into practical applications through partnerships with industry stakeholders, which will enable us to test its effectiveness in real-world scenarios and refine it further based on practical feedback. This iterative refinement and application will help in tailoring the algorithm to better meet specific operational needs, ultimately enhancing its usability and impact..

*IX.*REFERENCES

[1] Mona KÃ¶hler,Markus Eisenbach,Horst-Michael Gross Few-Shot Object Detection: A Comprehensive Survey IEEE Transactions on Neural Networks and Learning Systems, 2023

[2] Yang Xu,Bohao Huang,Xiong Luo,Kyle Bradbury,Jordan M. Malof SIMPL: Generating Synthetic Overhead Imagery to Address Custom Zero-Shot and Few-Shot Detection Problems IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022

[3] Yuchen Zhang,Bo Zhang,Bin Wang Few-Shot Object Detection With Self-Adaptive Global Similarity and TwoWay Foreground Stimulator in Remote Sensing Images IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022

[4] Zixuan Xiao,Jiahao Qi,Wei Xue,Ping Zhong Few-Shot Object Detection With Self-Adaptive Attention Network for Remote Sensing Images IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2021

[5] Zhaoyang Wang,Qiang Gao,Dong Li,Junjie Liu,Hongwei Wang,Xiao Yu,Yipin Wang Insulator Anomaly Detection Method Based on Few-Shot Learning IEEE Access, 2021

[6] S. M. Gubanski, A. Dernfalk, J. Andersson and H. Hillborg, Diagnostic methods for outdoor polymeric insulators, IEEE Trans. Dielectr. Electr. Insul., vol. 14, no. 5, pp. 1065-1080, Oct. 2007.

[7] N. C. Mavrikakis, P. N. Mikropoulos and K. Siderakis, Evaluation of field-ageing effects on insulating materials of composite suspension insulators, IEEE Trans. . .

[8] B. Wang, M. Ren, Z. Wu, C. Guo, T. Zhuang, et al., Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis, IEEE Trans. Instrum. Meas., vol. 69, no. 8, pp. 5345-5355, Aug. 2020.

[9] N. Carion, End-to-end object detection with transformers, Proc. Eur. Conf. Comput. Vis., pp. 213-229, 2020.

[10]Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, Proc. Int. Conf. Neural Inf. Process. Syst., pp. 1097-1105, 2012.

[11] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770-778, Jun. 2016.

Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1-9, Jun. 2015.