Recursive Gated Convolution-Based YOLOv11 Framework for Operator Safety Management in Live-Line Work

Dapeng Ma, Liang Yang, Kang Chen, Feng Yang, Ao Cui, Rundong Yang*, Zhilin Wen, Donghua Zhao
Yunnan Power Grid Co., Ltd., Kunming, China

***Abstract*—In live-line work scenarios, it is essential for workers to wear electric field shielding clothing to prevent fatal accidents caused by electric shock. Accordingly, this study developed an electric field shielding clothing detection system for live-line working environments based on the YOLOv11 framework. Previous research has explored intelligent wearable detection systems for personal protective equipment such as safety helmets. However, compared to safety helmets, electric field shielding clothing comes in more varieties and is more challenging to identify. To address the challenges mentioned above, this study constructed a dual-layer detection model for operator detection and electric field shielding clothing detection in live-line work scenarios. The first layer employs an improved detection transformer (IDETR) to locate operators within the environment. The second layer, based on the YOLOv11 framework integrated with recursive gated convolution (GnConv), is designed to classify three types of personal protective equipment, including electric field shield clothing, electric field shield masks, and electric field shield gloves. Finally, the experimental results showed that compared with the DETR, the accuracy of the IDETR-based worker localization model improved by 2.29%. The accuracy of the GnConv-based YOLOv11 framework in the electric field shielding clothing detection task reaches 90.40%.**

***Keywords*—Detection transformer; recursive gated convolution; YOLOv11; personal protective equipment; live-line work scenarios**

I. INTRODUCTION

Electric field shielding clothing (EFSC) protects workers from high-voltage electric fields, induced currents, and potential injuries caused by contact currents. In live-line working environments, failure to use EFSC correctly or at all can expose personnel to life-threatening risks. Therefore, the development of a real-time monitoring system for EFSC can effectively prevent workers who are not wearing or improperly wearing the protective gear from entering hazardous areas or performing high-risk operations [1]. In this study, the real-time detection system for EFSC involves using visual cameras and image recognition technology to detect whether workers in live line work scenes are wearing EFSC in a standardized manner. The system primarily consists of two modules: the staff positioning and the EFSC wearing detection module. For the staff positioning module, after the visual camera captures images of the staff, this module functions to accurately identify both personnel and the working environment. The EFSC wearing detection module utilizes the output from the positioning module to detect specific types of personal protective equipment, including electric field shield clothing, masks, and gloves.

In the context of live-line maintenance, accurate detection of personnel presents several challenges. First, complex environmental backgrounds, such as interference from transmission towers and other equipment, adversely affect target localization. Second, the workers themselves, as the targets of detection, exhibit considerable complexity. Their postures are non-standard and highly dynamic, and they may occlude one another [2]. The detection of EFSC is critical to ensuring the safety of live-line work, yet this task also involves numerous technical challenges in practical applications. Firstly, the material of the EFSC has a metallic texture and typically appears in silver or black. These colors are prone to strong reflections or shadows under certain lighting conditions, which can blur visual features such as the edges of the clothing. This directly increases the difficulty of image segmentation and feature extraction in the EFSC wearing detection module. Moreover, some critical safety items, such as electric field shielding gloves and protective eyewear, occupy only a small area within captured images. Their subtle features are easily overlooked during object recognition, leading to missed detections or misidentification. Additionally, the cost of acquiring real-world live-line maintenance images is high, and it is challenging to accurately annotate boundaries of occluded or reflective areas on EFSC in the images. This issue directly impacts the performance of model training [3].

Currently, mainstream object detection methods include transformer-based and convolutional neural network (CNN)-based approaches. CNN-based methods for staff localization require pre-defined anchor boxes of various sizes and aspect ratios for all images in both the training and validation sets, resulting in a more complex workflow. In contrast, transformer-based object detection methods significantly reduce the need for manual prior knowledge and tedious image preprocessing [4]. In addition, compared to CNN-based methods for worker localization, the transformer-based model demonstrates superior performance in detecting occluded objects [5]. Therefore, to address the object detection task in live-line work scenarios, this study introduces an improved version of the detection transformer (DETR) model, with the aim of enhancing the detection of transmission line maintenance workers in complex environments. However, the DETR model also requires a substantial number of training epochs to converge and exhibits poor performance in small object detection due to its high sampling rate [6]. Therefore, for

*Corresponding author.

the task of classifying three types of personal protective equipment, electric field shielding clothing, electric field shielding masks, and electric field shielding gloves, an improved YOLOv11 framework was developed.

In summary, this study proposes a dual-layer detection model to enhance the safety management of operators during live-line work. The first layer consists of a staff positioning module based on an improved DETR model, while the second layer comprises a personal protective equipment detection module based on an enhanced YOLOv11 architecture tailored for electric field shielding gear. Specifically, a deformable attention (DA) module has been incorporated into the DETR model to improve both its convergence speed and object detection accuracy. Additionally, the recursive gated convolution (GnConv) module has been integrated into the YOLOv11 framework to boost its performance in detecting small objects. The overall structure of the proposed dual-layer safety management detection model for transmission line maintenance operators is illustrated in Fig. 1. The main contributions of this study are summarized as follows:

- A dual-layer detection model for safety management of workers in live-line working scenarios has been proposed, which effectively integrates the tasks of identifying electric field shielding equipment and locating operators.
- The DA module is introduced into the DETR model to improve the accuracy of target detection for operators in live-line working scenarios.
- The GnConv module is embedded into the YOLOv11 architecture, aiming to improve the feature extraction and classification capabilities of the YOLO architecture for small target objects such as electric field shielding clothing.
- Experiments were conducted based on image information collected from real live-line working scenarios, and the results showed that the improved YOLOv11 framework achieved a detection accuracy of 92.61% in the electric field shielding clothing detection task.

The remaining sections of this study are arranged as follows: Section II reviews the work related to personal protective equipment testing. Section III describes the proposed object detection framework. Section IV presents the results and discussion. Finally, Section V provides a summary of the entire study.

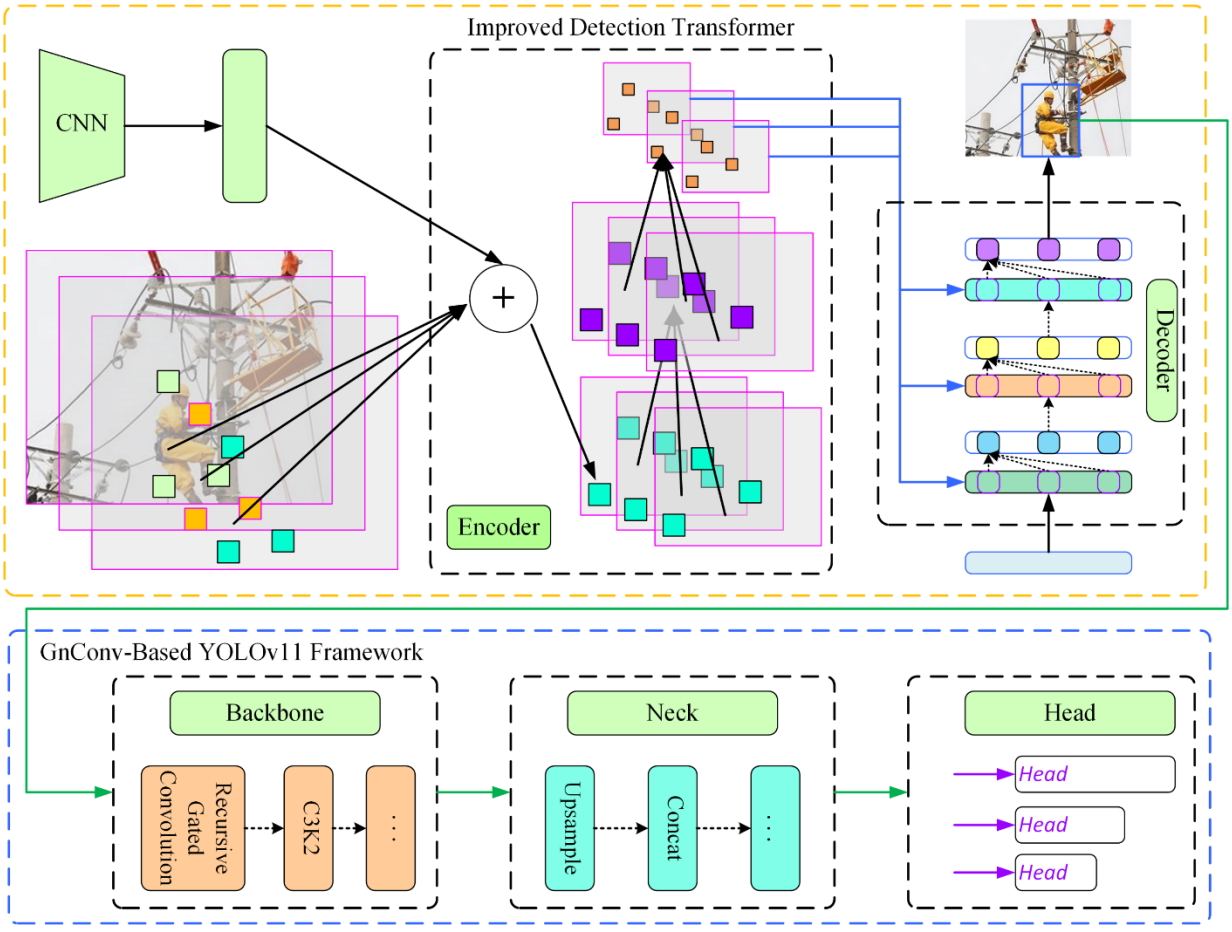

Fig. 1. The proposed dual-layer detection model for operator safety management in live-line work.

II. Related Work

The double-layer target detection model proposed in this study, designed for the safety management of workers in live-working scenarios, comprises two main tasks: worker positioning and electric field shielding equipment identification. Accordingly, a review of related work has been conducted, covering transformer-based image processing models and object detection frameworks based on YOLO.

A. Detection Transformer

In [7], the authors reviewed transformer-based image processing models. Specifically, it includes the basic architecture, core mechanisms, and key improvements and variations of transformer-based image processing models in different image processing tasks. In addition, this study comprehensively summarizes the benefits of transformers compared to traditional CNNs in processing visual data. In [8], the authors improved the DETR model and designed a lightweight transformer-based object detection model. In this study, an encoder-free neck (EFN) architecture was designed to reduce the computational overhead of traditional DETR models during training. The experimental results show that the proposed lightweight DETR model improves operational efficiency through structural optimization while maintaining end-to-end detection advantages, achieving a balance between high efficiency and detection accuracy.

In [9], the authors improved the DETR object detection framework to handle object detection tasks in pathological images. In response to the challenges of small cell scales and dense distribution in organizational images, this study optimized the feature extraction and matching mechanism of DETR, significantly improving the recall and robustness of mitotic cell recognition. Choi et al. focused on object detection in dense scenes and proposed a DETR model based on recurrent [10]. This model enhances DETR's ability to distinguish occluded targets by introducing a recursive mechanism, and experiments were conducted on a typical dense target dataset. The results showed that the DETR model based on recurrent improved the detection accuracy of traditional DETR models in high occlusion situations, providing a foundation for the improvement of DETR models.

Ghahremani et al. proposed a Transformer-based organ detection method aimed at improving the accuracy and robustness of organ localization in medical images [11]. This model introduces a deformable attention mechanism in the DETR framework to enhance the feature capture ability of classical DETR models for organs of different sizes and shapes, solving the problem of low target localization accuracy in traditional CNN-based detection methods in complex medical images. In [12], the authors combine the Mamba model with transformer for small object detection tasks, aiming to leverage the efficiency advantage of Mamba in long sequence modeling and enhance the ability to extract multi-scale features of small objects. The final results indicate that the Mamba-based transformer model significantly improves detection accuracy while maintaining low computational overhead. In [13], the authors proposed an end-to-end rotating object detection transformer framework. This framework overcomes the challenge of traditional horizontal detection boxes being sensitive to directional changes by designing an angle aware query mechanism and a rotation box alignment loss function.

In [14], the authors also improved the DETR model, aiming to enhance the detector's discriminative ability in complex backgrounds. The improved DETR model can more accurately distinguish aircraft targets from background interference by introducing a loss function of structural perception. In [15], the authors studied the synchronization loss optimization problem in rotation and orientation object detection based on the DETR model. This study proposes a universal synchronization optimization strategy to improve training stability. The experimental results show that the proposed method effectively alleviates the conflict between angle regression and classification tasks by unifying different loss calculation methods.

B. YOLO-Based Target Detection

In [16], the authors explore the application of YOLOv11 in urban map drawing. This study introduces the advanced real-time object detection model YOLOv11 into complex urban environments, aiming to enhance the automatic recognition and annotation capabilities of urban features. Perikamana Narayanan et al. designed a face detection and counting system based on YOLOv9, focusing on the confusion between human and animal faces and the interference of complex imaging environments on the accuracy of object detection models [17]. Finally, the experimental results demonstrate that the YOLOv9 model can maintain high recognition accuracy and robustness under complex conditions. Alsabei et al. applied YOLOv9 to the field of security, aiming to solve the real-time detection task of abnormal pedestrian behavior [18]. This study focuses on the characteristics of dense crowds and complex behavioral patterns in high-risk scenarios, and uses YOLOv9 to detect abnormal activities that may cause danger in real time. Khan et al. utilized an optimized YOLOv9s model for real-time road damage detection [19]. This study enhanced the performance of YOLOv9s through optimization and fusion techniques of object detection models, thereby achieving real-time detection of defects such as road surface damage and cracks.

In [20], the authors proposed a small object detection model in remote sensing images based on YOLOv10, aiming to solve the problem of low resolution and weak features of small objects in remote sensing images. This study significantly improved the recognition accuracy and recall rate of YOLOv10 for small objects in complex remote sensing scenes by improving it. In [21], the authors propose an intelligent psyllid monitoring system based on YOLOv10, which combines the YOLOv10 model with a visual transformer for detecting small pest targets that are difficult to detect in agricultural scenes. In [22], the authors proposed a deep-sea fish detection model based on YOLOv10. This model has the advantages of lightweight and high detection accuracy. Overall, the model developed in this study effectively addresses challenges such as dim lighting, complex backgrounds, and diverse forms of fish targets in deep-sea exploration. In [23], the authors developed a traffic police gesture recognition framework based on the YOLO model. Based on the latest YOLOv11 framework, the YOLOv11 framework was improved to address the challenges of fast

dynamic changes in traffic police gesture actions and high background noise interference in traffic command scenarios.

In [24], the authors developed a method combining faster region-based convolutional neural networks (Faster R-CNN) with model driven clustering for special object detection problems with large aspect ratios. This study effectively addresses the challenge of poor detection performance of general object detection models for unconventional shaped objects. In [25], the authors integrate real-time DETR (RT-DETR) and ByteTrack algorithm to construct a multi vehicle tracking and counting framework for daily traffic flow survey. This study demonstrates the application effectiveness of Transformer based detectors in practical engineering. In [26], the authors developed a precise reading algorithm for substation pointer instruments based on RetinaNet. By improving the RetinaNet detection network, it effectively solved the reading difficulties caused by uneven lighting, dial fouling, and small pointer shapes in actual industrial environments.

III. The Proposed Target Detection Model

The proposed live-line work scenario operator security management model mainly consists of two components. The first component involves the use of the deformable attention-based DETR (DA-DETR) for the operator localization task. The second component integrates GnConv and YOLOv11, aiming to detect whether workers are wearing electric field shielding clothing according to standards.

A. The Deformable Attention-based DETR

Operator localization in live working scenarios is a challenging object detection task. Specifically, the background of live working sites is complex, with interferences such as wires, insulators, and towers. In addition, the varying postures of operators result in significant differences in target scale. To address the above challenges, the DA mechanism has been introduced into the DETR model, aiming to improve the accuracy of object detection in classical DETR models. Fig. 2 illustrates the DA mechanism. The input data for the operator positioning system based on DA-DETR is usually real-time video streams collected by visual sensors from live work sites. In the encoding stage, the Transformer encoder utilizes the DA mechanism to enhance and extract the global features of live working site images. In the decoding stage, a set of learnable object queries interact with image features through the DA mechanism, enabling them to adaptively focus on the most informative key regions and accurately locate potential targets.

In the DA-DETR-based operator positioning system, the DA mechanism uses a multi-head attention model and a deformable attention model. The definition of the multi-head attention model is as follows [see Eq. (1)]:

$$MHAtten(\Pi_i, D_e) = \sum_{h \in H} W_h \times \left[\sum_{e \in E} A_{h,i,e} \times W_h^* \times D_e \right] \quad (1)$$

where, Π_i is the i -th element of the feature matrix. D_e is the e -th vector of the input data. $h \in H$ is the attention head. W_h and W_h^* are learnable weight parameters. $A_{h,i,e}$ is the attention weight.

The constraint conditions of $A_{h,i,e}$ are defined as Eq. (2):

$$\sum A_{h,i,e} = 1 \quad (2)$$

For a given feature map D_e , generate a point B_e as a unified grid reference. The definition of deformable attention model is as follows [see Eq. (3)]:

$$DAtten(\Pi_i, D_e) = \sum_{h \in H} W_h \times \left[\sum_{e \in E} A_{h,i,e} \times W_h^* \times (B_e + \Delta B_{h,i,e}) \right] \quad (3)$$

where, $\Delta B_{h,i,e}$ is the offset of point B_e .

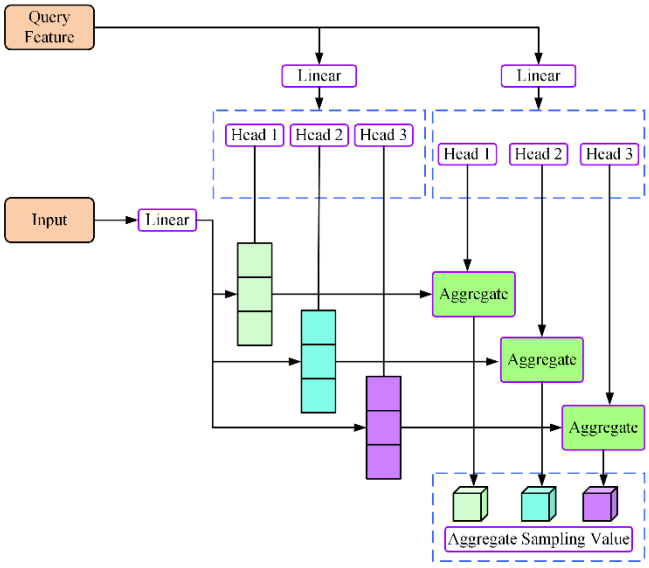

Fig. 2. The deformable attention (DA) mechanism.

B. The GnConv-Based YOLOv11

The electric field shielding clothing wearing detection system uses a hybrid GnConv and YOLOv11 (GnConv-YOLOv11) framework, which embeds the GnConv module into the architecture of YOLOv11. The advantage of GnConv lies in its ability to establish global context dependencies based on super large convolutional kernels, enabling CNN to detect the global nature of images. The GnConv-YOLOv11 framework can capture the subtle features of small targets such as electric field shielding goggles and gloves, as well as their correlation with live working environments, significantly improving the ability to discover and identify small targets in the context of live-line working. After the DA-DETR model locates the workers, the electric field shielding clothing wearing detection system uses a backbone convolutional network embedded with GnConv to extract the features of the electric field shielding clothing. Then, the neck network integrates multi-scale features, and finally the detection head outputs accurate classification results. Fig. 3 shows the structure of GnConv.

The core operation of GnConv is gated Convolution, which is defined as $GOConv(.)$ in this study. When the input feature is $L \in \mathbb{R}^{C1 \times C2}$, the output y of $GOConv(.)$ operation is defined as Eq. (4):

$$\hat{y} = \phi_{output}(U_1) \in \mathbb{R}^{C1\times C2} \quad (4)$$

where, $\phi_{output}(.)$ is a linear projection operation. U_1 is the convolution result of the depth wise operation of the electric field shielding clothing wearing detection system.

U_1 is defined in Eq. (5):

$$U_1 = f(R_0) \odot U_0 \in \mathbb{R}^{C1\times C2} \quad (5)$$

where, $f(.)$ is the depth-wise convolution operation. R_0 and U_0 are the features of input $L \in \mathbb{R}^{C1\times C2}$, and the calculation equation is as follows [see Eq. (6)]:

$$\left[R_0 \in \mathbb{R}^{C1\times C2}, U_0 \in \mathbb{R}^{C1\times C2}\right] = \phi_{in}(L) \in \mathbb{R}^{C1\times 2C2} \quad (6)$$

where, $\phi_{in}(.)$ is a linear projection operation.

Therefore, the definition of $GOConv(.)$ operation is Eq. (7):

$$\hat{y} = GOConv\left(L \in \mathbb{R}^{C1\times C2}\right) \quad (7)$$

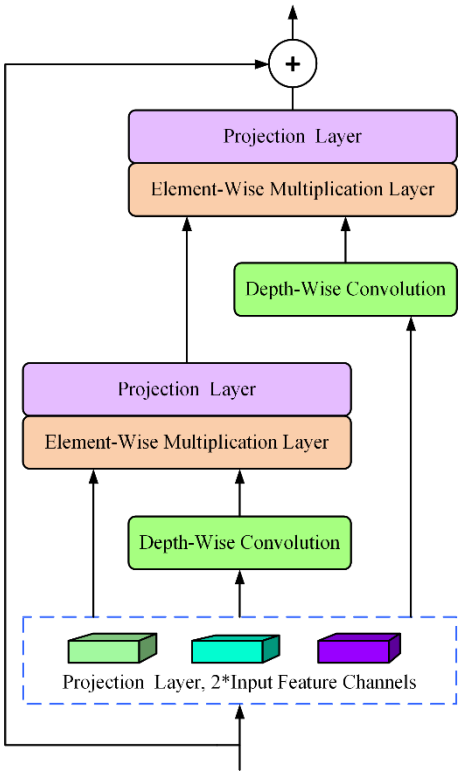

Fig. 3. The structure of recursive gated convolution (GnConv).

IV. Results and Discussion

The DA-DETR model is used for operator object detection in live-line work scenarios. For the operator object detection model, the operator image in the live-line work scenario is used as input. 9100 images of operators in live-line work scenarios were used as samples for the training and validation sets. The epochs of the DA-DETR based operator object detection model are set to 24, with 1000 steps per epoch. During the training process, when the confidence level of the identified personnel within the bounding box is greater than 70%, it is defined as a successful detection of the personnel. The output of the operator object detection model, the range of the bounding box, is used as the input for the electric field shielding clothing wearing detection system. The electric field shielding clothing wearing detection model aims to detect three types of electric field shielding equipment, as shown in Fig. 4. Table I introduces three types of electric field shielding equipment.

Fig. 4. Personal protective equipment in live-line work scenarios.

TABLE I. Three Types of Electric Field Shielding Equipment

Illustration	Name	Category
Fig.4. (a), (b), and (c)	Electric field shield clothing	3
Fig.4. (d), (e), and (f)	Electric field shield gloves	3
Fig.4. (g), (h), and (i)	Electric field shield mask	3

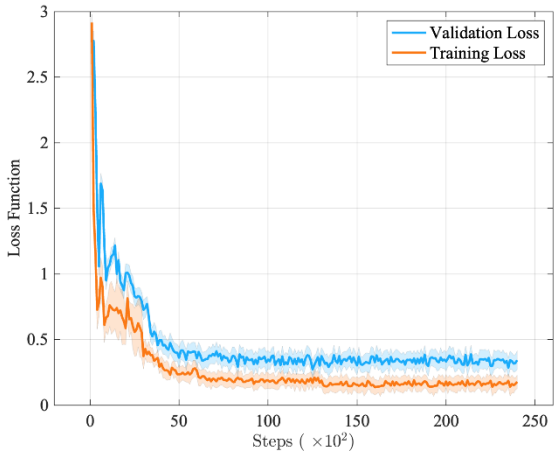

Fig. 5. The loss function curve of operator positioning model based on DA-DETR.

Due to the fact that the training process of machine learning is essentially random, in order to verify the robustness of the

DA-DETR model and GnConv-YOLOv11 model, each model was trained 15 times, separately. The loss functions of the DA-DETR model and GnConv-YOLOv11 model during 15 training sessions are shown in Table II.

TABLE II. THE LOSS FUNCTIONS OF THE DA-DETR MODEL AND GNCONV-YOLOV11 MODEL

Loss function	Training loss		Validation loss	
	DA-DETR	***GnConv-YOLOv11***	***DA-DETR***	***GnConv-YOLOv11***
Average	0.1762	0.1316	0.3415	0.2151
The upper bound of 95% C.I.	0.2278	0.1987	0.4130	0.2909
The lower bound of 95% C.I.	0.1246	0.0643	0.2700	0.1349

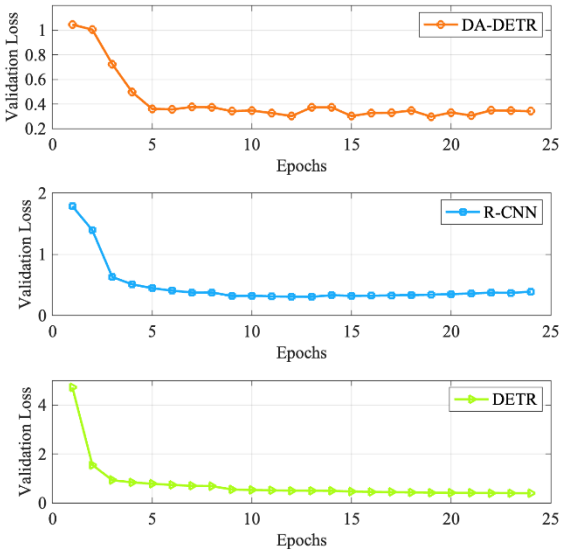

Fig. 6. The loss function values of the DA-DETR model recorded during 24 epochs of training.

Fig. 5 shows the loss function of the operator localization model based on DA-DETR. Fig. 6 shows the average loss function of the operator localization model based on DA-DETR over 24 iteration cycles. To demonstrate the effectiveness of the DA-DETR model in handling operator localization problems, a comparison was made between the DA-DETR model, R-CNN [24], and DETR model. Fig. 6 also shows the comparison results.

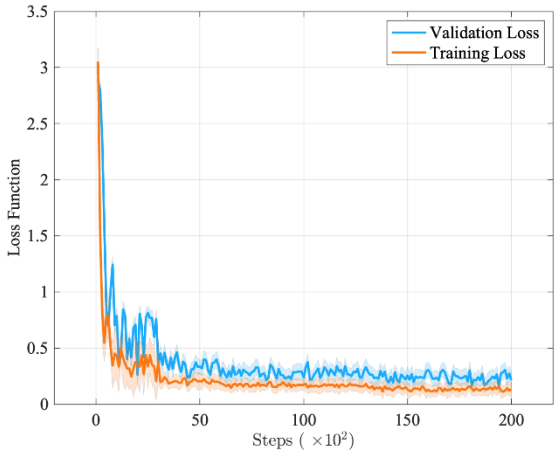

Fig. 7. The loss function curve of the GnConv-based YOLOv11.

Table III shows the results of DA-DETR, YOLOv9, YOLOv7, Fast R-CNN, R-CNN, and DETR [14] models in handling operator localization tasks in the context of live working. Each model is evaluated from six perspectives, including training loss, validation loss, recall rate, precision, accuracy, and F1-Score. The training loss of the operator localization model based on DA-DETR is 0.1763. The DA-DETR model has the lowest loss, with a 2.54% reduction in training loss compared to the YOLOv9 model. The validation loss of the operator positioning model based on DA-DETR is 0.3415, which is 3.53% lower than that of the YOLOv9 model. This indicates that the DA-DETR-based operator localization model learns most effectively from the training data, with its internal parameters efficiently capturing features from images of operators in live-line work scenarios. Moreover, the DA-DETR model's lower validation loss compared to all other models demonstrates that it possesses the strongest generalization capability. For the precision indicator, the accuracy of the operator positioning model based on DA-DETR is 0.9083, while the best-performing comparison model, YOLOv7, is 0.9013. Compared with the YOLOv7 model, DA-DETR achieved a performance improvement of 0.78% in accuracy. This improvement indicates that DA-DETR has a slight advantage in reducing false positives for operators in the context of live working, and the prediction results of DA-DETR are more reliable.

TABLE III. THE RESULTS OF OPERATOR POSITIONING MODELS

Algorithm	Training Loss	Validation Loss	Recall	Precision	Accuracy	F1-Score
DA-DETR	0.1763	0.3415	0.9194	0.9083	0.9118	0.9118
YOLOv9	0.1809	0.3540	0.9080	0.9020	0.9050	0.9050
YOLOv7	0.1952	0.3785	0.8998	0.9013	0.9043	0.9043
Fast R-CNN	0.1829	0.3617	0.9050	0.8976	0.9013	0.9013
R-CNN	0.1967	0.3897	0.8955	0.8871	0.8913	0.8913
DETR	0.2132	0.3943	0.8902	0.8917	0.8909	0.8909

Table III also shows the recall rate, accuracy, and F1-score. In this study, the recall rate is the ratio of successful predictions made by the operator localization model among all real operator samples. A high recall rate means that the operator has a low rate of missed detections in the localization model. The recall rate of the operator positioning model based on DA-DETR is 0.9194, and the recall rate of the second-ranked YOLOv9 model is 0.9080. Compared with the YOLOv9 model, the recall rate of the DA-DETR model has increased by 1.25%. Therefore, DA-DETR has the highest recall rate when handling operator positioning tasks in the context of live working, indicating its strongest ability to identify operators and find more operators in the image. The F1-score of DA-DETR is 0.9118, and DA-DETR also achieved the highest F1-score. This proves that the DA-DETR operator localization model achieves the optimal balance between accuracy and recall, achieving the best overall performance. Finally, the accuracy of the DA-DETR model is 0.9118, which is 0.75% higher than the YOLOv9 model. This indicates that the DA-DETR operator localization model has achieved lower missed detection rates (high recall rates) and lower false detection rates (high accuracy rates). Therefore, the operator positioning model based on DA-DETR can handle complex live working backgrounds and effectively prevent misjudging transmission lines, insulators, and trees as workers.

Fig. 7 shows the iterative loss function curve of the GnConv-YOLOv11-based electric field shielding clothing wearing detection model. Table IV shows the application results of the GnConv-YOLOv11 model, YOLOv11 [16] model, YOLOv10 model, YOLOv9 [17] model, YOLOv8 model, YOLOv7 model, RT-DETRv2 model, RT-DETR [25] model, and RetinaNet [26] model in the problem of electric field shielding clothing wearing detection. Fig. 8 shows the recognition results of some electric field shielding clothing detection models.

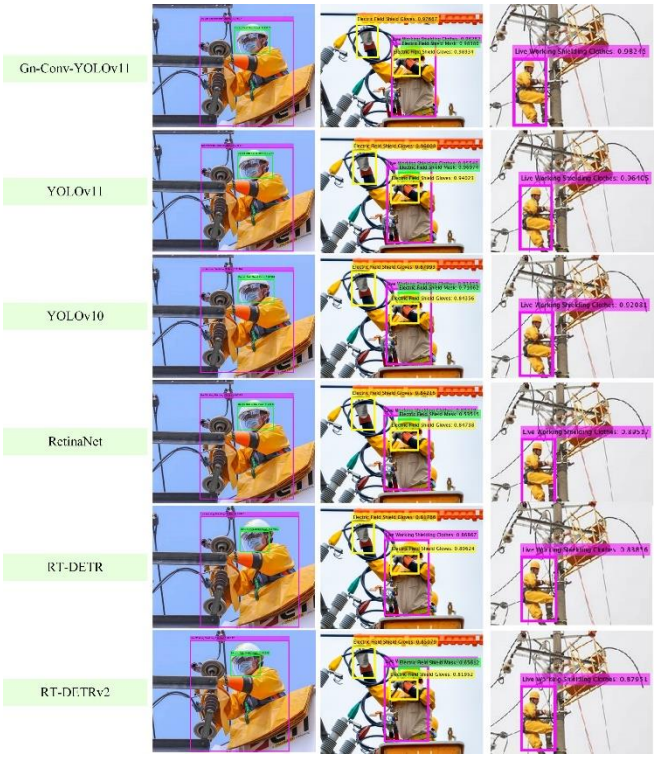

Fig. 8. Identification results of electric field shielding clothing detection models.

TABLE IV. THE RESULTS OF THE ELECTRIC FIELD SHIELDING CLOTHING DETECTION MODELS

Algorithm	Training Loss	Validation Loss	Macro-Recall	Macro-Precision	Accuracy
GnConv-YOLOv11	0.1316	0.2151	0.9157	0.8963	0.9040
YOLOv11	0.1795	0.2676	0.9019	0.8780	0.8931
YOLOv10	0.1978	0.2975	0.8854	0.8840	0.8831
YOLOv9	0.2143	0.3213	0.8784	0.8770	0.8777
YOLOv8	0.2382	0.3346	0.8623	0.8610	0.8584
YOLOv7	0.2479	0.3564	0.8556	0.8544	0.8550
RT-DETRv2	0.1872	0.2996	0.8816	0.8802	0.8812
RT-DETR	0.2214	0.3314	0.8646	0.8632	0.8639
RetinaNet	0.2526	0.3587	0.8512	0.8498	0.8505

In Table IV, five metrics including training loss, validation loss, macro-recall, macro-precision, and accuracy were used to evaluate all electric field shielding clothing wearing detection models. Firstly, the training loss of the GnConv-YOLOv11 model is 0.1316, which is lower than the other 8 models and reduces the training loss by about 26.7% compared to the benchmark YOLOv11. This indicates that the newly introduced GnConv-YOLOv11 model greatly enhances the learning ability of the electric field shielding clothing wearing detection model, enabling it to extract features more accurately from training data. The validation loss of the GnConv-YOLOv11-based electric field shielding clothing wearing detection model is 0.2151, which is about 19.6% lower than the second ranked YOLOv11 model. The GnConv-YOLOv11 model achieved an accuracy of 0.9040, which is higher than that of all eight other compared models for detecting the wearing of electric field shielding clothing, representing a 1.22% improvement over YOLOv11. Furthermore, the GnConv-YOLOv11-based model for detecting the wearing of electric field shielding clothing also achieved optimal performance in both macro-precision and macro-recall. Specifically, it exhibited a 2.08% improvement in macro-

precision and a 1.53% increase in macro-recall compared to the YOLOv11 model.

In addition, compared to the other eight electric field shielding clothing detection models, the GnConv-YOLOv11 model achieved an average reduction of 36.5% in training loss and 30.4% in validation loss. Meanwhile, it improved macro recall, macro precision, and accuracy by an average of 4.11%, 2.83%, and 3.44%, respectively. These results demonstrate that GnConv-YOLOv11 achieves comprehensive performance improvements across all metrics when compared to the other eight algorithms. It is also worth noting that the RT-DETR model performs similarly to YOLOv9 and YOLOv10, suggesting that CNN-based architectures remain highly competitive in the field of small object detection.

V. Conclusion

This study designed a dual-layer detection system for intelligent detection of electric field shielding clothing to meet the high-risk operational requirements in live-line work scenarios in the power industry. The system achieves accurate recognition of workers and their electric field shielding clothing by constructing an operator positioning model based on DA-DETR and a classification model that integrates GnConv and YOLOv11. Experimental results have shown that the DE-DETR model improves the accuracy of worker localization tasks by 2.29% compared to traditional DETR models. The electric field shielding clothing classification model based on GnConv-YOLOv11 achieved a detection accuracy of 92.61%, significantly better than mainstream detection models such as the YOLOv11 model, RT-DETR model, and RetinaNet model.

In addition, GnConv-YOLOv11 achieves the best performance across four key aspects: validation loss, macro recall, macro precision, and accuracy. Compared to the average of all benchmark models, it shows significant improvements in five major metrics: a 36.5% reduction in training loss, a 30.4% reduction in validation loss, a 4.11% increase in macro recall, a 2.83% improvement in macro precision, and a 3.44% gain in accuracy. Therefore, the proposed double-layer detection framework effectively tackles the challenges posed by the high diversity and difficulty of recognizing electric field shielding clothing in live-line work environments, offering a reliable technical basis for intelligent safety monitoring in such scenarios. The system proposed in this study has not been deployed in the real world. In future work, the developed electric field shielding clothing detection system will be deployed on unmanned aerial vehicles (UAVs) for real-world applications.

Acknowledgment

This work was supported by the Science and technology project of China Southern Power Grid Co., Ltd. under Grants YNKJXM20240219 and YNKJXM20240472.

References

[1] M. Zhao and M. Barati, "Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection Using GNN Approach," IEEE Transactions on Industry Applications, vol. 59, no. 3, pp. 3142-3150, May-June 2023, doi: 10.1109/TIA.2023.3234515.

[2] P. Illy and G. Kaddoum, "A Collaborative DNN-Based Low-Latency IDPS for Mission-Critical Smart Factory Networks," IEEE Access, vol. 11, pp. 96317-96329, 2023, doi: 10.1109/ACCESS.2023.3311822.

[3] H. Zheng, S. Xu, J. Li, F. Gao and Z. Cui, "A Lightweight Method Integrating Keypoint Detection and Perspective Geometry for Substation Safety Distance Monitoring," IEEE Transactions on Power Delivery, vol. 40, no. 2, pp. 810-821, April 2025, doi: 10.1109/TPWRD.2024.3522808.

[4] Sami Aziz Alshammari, "TLDViT: A Vision Transformer Model for Tomato Leaf Disease Classification" International Journal of Advanced Computer Science and Applications (IJACSA), 15 (12), 2024. http://dx.doi.org/10.14569/IJACSA.2024.0151285

[5] M. Alshehri, A. Ouadou and G. J. Scott, "Deep Transformer-Based Network Deforestation Detection in the Brazilian Amazon Using Sentinel-2 Imagery," IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1-5, 2024, Art no. 2502705, doi: 10.1109/LGRS.2024.3355104.

[6] S. Diop, F. Jouen, J. Bergounioux and I. Trabelsi, "Fine-Tuned YOLO Model for Monitoring Children Across Medical Scenes Based on a Large-Scale Real-World Dataset for Children Detection," IEEE Access, vol. 13, pp. 130953-130962, 2025, doi: 10.1109/ACCESS.2025.3588316.

[7] Ch. Sita Kameswari, Kavitha J, T. Srinivas Reddy, Balaswamy Chinthaguntla, Senthil Kumar Jagatheesaperumal, Silvia Gaftandzhieva and Rositsa Doneva, "An Overview of Vision Transformers for Image Processing: A Survey" International Journal of Advanced Computer Science and Applications (IJACSA), 14 (8), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140830

[8] Siyi Cheng, Jingnan Song, Mingliang Zhou, Xuekai Wei, Huayan Pu, and Jun Luo, "EF-DETR: A Lightweight Transformer-Based Object Detector With an Encoder-Free Neck," IEEE Transactions on Industrial Informatics, vol. 20, no. 11, pp. 12994-13002, Nov. 2024, doi: 10.1109/TII.2024.3431044.

[9] F. B. K. Ardaç and P. Erdogmus, "Mi-DETR: For Mitosis Detection From Breast Histopathology Images an Improved DETR," IEEE Access, vol. 12, pp. 179235-179251, 2024, doi: 10.1109/ACCESS.2024.3492275.

[10] H. K. Choi, C. K. Paik, H. W. Ko, M. -C. Park and H. J. Kim, "Recurrent DETR: Transformer-Based Object Detection for Crowded Scenes," IEEE Access, vol. 11, pp. 78623-78643, 2023, doi: 10.1109/ACCESS.2023.3293532.

[11] M. Ghahremani, B. R. Ernhofer, J. Wang, M. Makowski and C. Wachinger, "Organ-DETR: Organ Detection via Transformers," IEEE Transactions on Medical Imaging, vol. 44, no. 6, pp. 2657-2671, June 2025, doi: 10.1109/TMI.2025.3543581.

[12] C. Li, Y. Hei, W. Li and Z. Xiao, "MIT-DETR: Mamba-in-Transformers for Efficient Small Object Detection in SAR Images," IEEE Signal Processing Letters, vol. 32, pp. 3097-3101, 2025, doi: 10.1109/LSP.2025.3582672.

[13] L. Dai, H. Liu, H. Tang, Z. Wu and P. Song, "AO2-DETR: Arbitrary-Oriented Object Detection Transformer," IEEE Transactions on Circuits and Systems for Video Technology, vol. 33, no. 5, pp. 2342-2356, May 2023, doi: 10.1109/TCSVT.2022.3222906.

[14] H. Liu, X. Ren, Y. Gan, Y. Chen and P. Lin, "DIMD-DETR: DDQ-DETR With Improved Metric Space for End-to-End Object Detector on Remote Sensing Aircrafts," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 4498-4509, 2025, doi: 10.1109/JSTARS.2025.3530141.

[15] G. Liu, Q. Yu, M. Gong and H. Yang, "Universal Synchronization Loss Optimization in DETR-Based Oriented and Rotated Object Detection," IEEE Access, vol. 13, pp. 45669-45681, 2025, doi: 10.1109/ACCESS.2025.3539874.

[16] Muhammad Emir Kusputra, Alesandra Zhegita Helga Prabowo, Kamel and Hady Pranoto, "Enhancing Urban Mapping in Indonesia with YOLOv11" International Journal of Advanced Computer Science and Applications (IJACSA), 16 (2), 2025. http://dx.doi.org/10.14569/IJACSA.2025.0160261

[17] S. Perikamana Narayanan, M. Sabarimalai Manikandan and L. R. Cenkeramaddi, "YOLOv9-Based Human Face Detection and Counting Under Human-Animal Faces, Complex Imaging Environments, and

Image Qualities," IEEE Access, vol. 13, pp. 129600-129637, 2025, doi: 10.1109/ACCESS.2025.3591247.

[18] A. A. Alsabei, T. M. Alsubait and H. H. Alhakami, "Enhancing Crowd Safety at Hajj: Real-Time Detection of Abnormal Behavior Using YOLOv9," IEEE Access, vol. 13, pp. 37748-37761, 2025, doi: 10.1109/ACCESS.2025.3545256.

[19] M. W. Khan, M. S. Obaidat, K. Mahmood, B. Sadoun, H. M. S. Badar and W. Gao, "Real-Time Road Damage Detection Using an Optimized YOLOv9s-Fusion in IoT Infrastructure," IEEE Internet of Things Journal, vol. 12, no. 11, pp. 17649-17660, 1 June1, 2025, doi: 10.1109/JIOT.2025.3537640.

[20] H. Sun, G. Yao, S. Zhu, L. Zhang, H. Xu and J. Kong, "SOD-YOLOv10: Small Object Detection in Remote Sensing Images Based on YOLOv10," IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 8000705, doi: 10.1109/LGRS.2025.3534786.

[21] Li Zhang, Qianyue Liang, Vijay John, Hong Chen, Shanjun Li, Weifu Li, "Intelligent Psyllid Monitoring Based on DiTs-YOLOv10-SOD," IEEE Transactions on AgriFood Electronics, vol. 3, no. 1, pp. 286-294, March-April 2025, doi: 10.1109/TAFE.2025.3551072.

[22] Z. Hu and Q. Chen, "MOA-YOLO: An Accurate, Real-Time, and Lightweight YOLOv10-Based Algorithm for Deep-Sea Fish Detection," IEEE Sensors Journal, vol. 25, no. 13, pp. 23933-23947, 1 July1, 2025, doi: 10.1109/JSEN.2025.3574723.

[23] Xuxing Qi, Cheng Xu, Yuxuan Liu, Nan Ma and Hongzhe Liu, “TPGR-YOLO: Improving the Traffic Police Gesture Recognition Method of YOLOv11” International Journal of Advanced Computer Science and Applications (ijacsa), 16 (2), 2025. http://dx.doi.org/10.14569/IJACSA.2025.0160243

[24] F. Fang, L. Li, H. Zhu and J. -H. Lim, "Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection," IEEE Transactions on Image Processing, vol. 29, pp. 2052-2065, 2020, doi: 10.1109/TIP.2019.2947792.

[25] Y. Gladiensyah Bihanda, C. Fatichah and A. Yuniarti, "Multi-Vehicle Tracking and Counting Framework in Average Daily Traffic Survey Using RT-DETR and ByteTrack," IEEE Access, vol. 12, pp. 121723-121737, 2024, doi: 10.1109/ACCESS.2024.3453249.

[26] X. Zhai, J. Tian and J. Li, "An Accurate Reading Algorithm for Substation Pointer Meters Based on Improved 2-D Gamma Function and PA-RetinaNet," in IEEE Sensors Journal, vol. 23, no. 12, pp. 13738-13750, 15 June15, 2023, doi: 10.1109/JSEN.2023.3274136.