IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 November 2023, accepted 13 December 2023, date of publication 19 December 2023,
date of current version 11 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3344805

RESEARCH ARTICLE

A Recurrent Approach for Uninterrupted Tracking of Rotor Blades Using Kalman Filter

YIMING XU[1], ZHENYU FU[1], WEI PENG[1], ZIHENG DING[1], GUAN LU[2], AND QIANG LIU[3]

[1]School of Electrical Engineering, Nantong University, Nantong 226019, China
[2]School of Mechanical Engineering, Nantong University, Nantong 226019, China
[3]School of Engineering Mathematics and Technology, University of Bristol, BS8 1QU Bristol, U.K.

Corresponding author: Guan Lu (luguan@ntu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61973178, in part by the Smart Grid Joint Fund of State Key Program of National Natural Science Foundation of China under Grant U2066203, in part by the Major Natural Science Projects of Colleges and Universities in Jiangsu Province under Grant 21KJA470006, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX22_3347.

ABSTRACT With the escalating requirements for maintenance of wind turbines, the deployment of Unmanned Aerial Vehicles (UAVs) for inspection tasks has become increasingly prevalent. However, wind turbine blades, which are thin and long, possess weak texture features that lead to target confusion when tracking specific parts of the dynamic blades. Additionally, wind turbine units, being large dynamic structures, often exceed the camera's field of view (FOV) and exhibit unique motion characteristics. These factors make the visual tracking of specific components unstable due to the lack of global motion information. In order to address the aforementioned challenges and achieve consistent calibration of key components under the dynamic operating conditions of wind turbines, this study has adopted a strategy of integrating the Squeeze-and-Excitation Network (SEnet) into the backbone network of YOLOv5. Innovatively, two hyperparameters have been introduced into the existing loss function to adjust the weights of samples under conditions of data imbalance, thereby enhancing the performance of the detection model. In the application of the DeepSORT tracking algorithm, Long Short-Term Memory (LSTM) networks have been combined to predict the trajectory of the rotor blade's central point, and an optimized Kalman filter has been employed to significantly improve the system's adaptability and precision under various motion conditions. Empirical results from this study underscore the efficacy of the proposed method, demonstrating its capability to accurately differentiate individual blades as well as specific blade segments. Compared to the traditional YOLOv5, the enhanced YOLOv5-SE has demonstrated a 5.3% improvement in the Mean Average Precision (mAP_0.5) evaluation metric. Moreover, the improved DeepSORT algorithm has exhibited high efficiency in maintaining continuous and stable tracking of moving blades, adeptly handling scenarios where rotor blades frequently enter and exit the FOV. This advancement paves the way for the broader application of UAVs in wind turbine inspections, offering the potential for more efficient and accurate maintenance protocols.

INDEX TERMS Target tracking, wind turbine, YOLOv5, motion estimation, unmanned inspection.

I. INTRODUCTION

As the scale of wind energy development rapidly expands, the field of inspection faces increasingly significant challenges. Traditional manual inspection methods, though feasible, have emerged with high costs and associated safety risks as primary constraints. Owing to their high efficiency, swift responsiveness, and exceptional reliability, UAVs are becoming the predominant choice for wind turbine unit inspections [1]. In non-stoppage scenarios, detecting the external condition of wind turbine generator units not only effectively enhances operational efficiency but also offers a crucial means to reduce maintenance costs. However, capturing the precise status of specific rotor blades during the

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Briso.

motion of wind turbines is a major challenge in intelligent inspections [2]. To address this, this study employs deep learning algorithms specifically designed for the detection and tracking of moving wind turbine generator unit rotor blades.

During wind turbine operations, the outer segments of their blades can reach linear speeds of over 90 m/s [3], leading to motion blur interference. The highly similar features among the three blades and the lack of distinct texture features on each blade make it challenging to identify specific segments. Moreover, with the span of wind turbine blades often reaching hundreds of meters [4], a drone's FOV during inspection missions can't encompass an entire blade. Consequently, the regions of interest frequently enter and exit the FOV as blades move. Ensuring consistent calibration of targets in long-term tracking is imperative for continuous and stable tracking of specific blade sections during observation.

In addressing the aforementioned challenges, this study conducted the following core work:

1. Optimization of DeepSORT for Rotor Blades: Considering the nonlinear motion characteristics exhibited by wind turbine rotor blades, DeepSORT faces certain tracking limitations. To mitigate this, we optimized the state variables and the state transition model of the Kalman filter. This ensured alignment with the nonlinear motion patterns of rotor blades, enhancing tracking accuracy and stability.

2. Utilization of LSTM for Uninterrupted Tracking: To ensure continuous stability in tracking the target region when rotor blades frequently enter and exit the FOV, we employed an LSTM model. This predicted the motion trajectory of rotor blades outside the camera's FOV, enabling the Kalman filter to continuously receive position information, thus achieving efficient and uninterrupted tracking.

3. Integration of SEnet with YOLOv5: In this research, we integrated the SEnet into the backbone network of YOLOv5. To further address the issue of class imbalance in the data, we adjusted the loss function and introduced two hyperparameters. This balanced the uneven distribution of samples and consequently improved detection accuracy.

Through experimental validations, our study has proven its capability to identify and track critical segments of specific blades under wind turbine motion conditions effectively. It ensures target consistency calibration of essential components of specific rotor blades in motion, enhancing the efficiency of intelligent drone inspections of wind turbine generator units.

II. RELATED WORKS

In recent years, computer vision technology applications have significantly expanded [5]. Research on the precision of target detection and the stability of target tracking has particularly gained traction, notably within the wind power sector, where the accurate detection and consistent tracking of wind turbine motion is crucial.

Advancements in detecting weak-textured feature targets have been significant. Ran et al. [6] and colleagues enhanced the real-time detection of weak texture defects on wind turbine blades by improving the feature pyramid network, reallocating input feature weights to better capture feature information, thus increasing the model's accuracy and robustness. Xiaoxun et al. [7] and his team focused on the efficient detection of surface cracks on wind turbine blades, particularly less visible light-colored cracks, by enhancing crack feature extraction under various lighting conditions using multi-source information and the C3TR module. Addressing the challenge of bolt loosening on wind turbines due to their non-distinct texture features, Yang et al. [8] and his team proposed a dual-phase detection framework that merges traditional manual torque techniques with deep learning models, enabling the recognition of bolts loosened by as little as 2 degrees under various conditions. Zhang and Wen [9] and his group introduced the SOD-YOLO model, optimized for rapidly detecting small target defects and other subtle imperfections. Yang et al. [10] and colleagues improved the detection precision of weak texture feature insulation defects on complex backgrounds by integrating Spatial Pyramid Pooling (SPP) with the MobileNet network. Xia et al. [11] and his team developed a unique reparameterized large-kernel C3 module specifically for weak-textured targets, combining adaptive receptive fields with multi-scale feature fusion to optimize detection of weak texture steel surface defects. Wang et al. [12] and his group created various feature extraction modules that combine depth, shape, and texture characteristics of detection targets, input into a cooperative network to enhance detection accuracy. Finally, Chen et al. [13] and his team developed a trapezoidal multi-attention network (LMNet) that excels in extracting features from weak-textured objects while minimizing feature information loss.

Nevertheless, the primary detection targets of this study, wind turbine blades, present highly similar features amongst each other, and distinct texture features within different sections of the blades are lacking, posing a challenge for recognizing specific blade sections.

Deep learning-based target tracking algorithms in the literature have shown substantial improvements in accuracy and robustness. Kim [14] achieved consistent calibration of high-maneuverability targets in 3D space by acquiring position and speed information of the tracking target and accounting for obstacles. Ning et al. [15] introduced STD-Yolov5, a new model embedding an attention mechanism into the primary network to enhance the network's feature extraction capacity, improving detection of weak texture targets against complex backgrounds. Wang and Huang [16] and his team used an improved gray neural network to track and pinpoint feature points of targets in videos, achieving consistent target calibration within video streams. Rao et al. [17] and associates proposed a multi-camera coordination strategy for motion target detection, tracking, and matching to meet the visual tracking requirements of specific targets. Du [18] and his team proposed a novel dynamic tracking approach for athletes based on wireless body area networks, capable

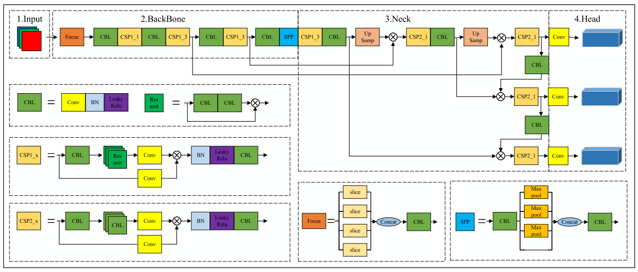


FIGURE 1. Schematic diagram of the YOLOv5 network structure.

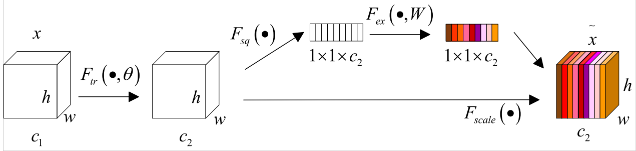


FIGURE 2. SEnet module structure.

of distinguishing and tracking similar targets for target consistency calibration.

Despite successes in the continuous and stable tracking of specific targets, wind turbines, the focus of this paper, often exceed the camera's FOV and display unique motion characteristics. The regions of interest frequently transition in and out of the FOV due to blade movement throughout the observation process, challenging the continuous and stable tracking of turbine blades and leading to unstable visual tracking of specific components because of the absence of global motion information.

In conclusion, while existing studies have made strides in detecting weak-textured targets and tracking them, significant challenges remain in addressing large dynamic structural objects like wind turbines with distinct motion characteristics against complex backgrounds. Thus, this study aims to bridge gaps in current research by integrating and optimizing the YOLOv5 backbone network, utilizing LSTM to predict the trajectory of the central point of the wind turbine blades, and enhancing the Kalman filter to improve system adaptability and accuracy, achieving consistent calibration of specific key components of the blades under operational conditions.

III. TECHNICAL SOLUTIONS

A. THE NETWORK STRUCTURE OF YOLOv5

This study utilizes YOLOv5 for the detection of wind turbine blade segments, the comprehensive design of which is depicted in Figure 1. The architecture of the YOLOv5 model can be broadly divided into four components: input, backbone, neck, and output [19]. As delineated in the figure, the input phase employs the Mosaic data augmentation technique. This method not only diversifies the background context of the detection target within the image but also bolsters the model's ability to detect smaller objects. Additionally, the input phase facilitates adaptive anchor box computations and dataset-specific image scaling. The backbone segment predominantly adopts the Focus and the Cross Stage Partial (CSP) structures. In the network's neck, integration of the Feature Pyramid Network (FPN) with the Personal Area Network (PAN) structure enhances multi-scale feature fusion capabilities. For the output phase, the model leverages a bounding box loss function, promoting faster and more efficient convergence.

B. IMPROVEMENT OF YOLOv5

1) INTEGRATING SEnet IN YOLOV5's BACKBONE NETWORK

As the depth of the YOLOv5 network increases, the extracted information at the output becomes progressively abstracted. Moreover, during the wind turbines' operation, the motion blur substantially affects the blade detection process. The considerable similarity among blade features, coupled with the absence of distinct texture patterns between blade sections, complicates the detection of specific blade components in drone inspection footage. In light of these challenges, this study undertakes refinements to the original YOLOv5. The modifications facilitate the learning of inter-channel feature relationships, amplifying the expression of pivotal channel features. This enhancement subsequently boosts the performance of the trained model.

The SENet is adept at discerning the significance of features across diverse channels, thereby enriching the feature map's representation of dimensional attributes. The architectural layout of the SENet is depicted in Figure 2.

For an input X with a shape of $C \times H \times W$, after passing through the convolutional layer, the feature map U is obtained. The calculation process is as follows:

$$F_{tr} : X \to U, X \in \Re^{H' \times W' \times C'}, \quad U \in \Re^{H \times W \times C}. \tag{1}$$

$$u_c = v_c * X = \sum_{s=1}^{c'} v_c^s * x^s. \tag{2}$$

The SEnet module is typically positioned after the convolutional layer and primarily consists of two operations:: Squeeze and Excitation. This module undertakes significance learning across various channels of U, thereby promoting the expression capability of the essential channels and concurrently attenuating the activity of the relatively weaker channels. Initially, the SENet module implements the Squeeze operation on U, deriving a one-dimensional feature z_c at the channel level. z_c stands as the preliminary weight coefficients for each respective channel. The procedure for the Squeeze operation is delineated in the subsequent equation:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{3}$$

In order to capture the activation intensity of feature representation for each channel, the initially obtained weight coefficients z_c undergo Excitation processing. The Excitation mechanism comprises two fully connected layers and two

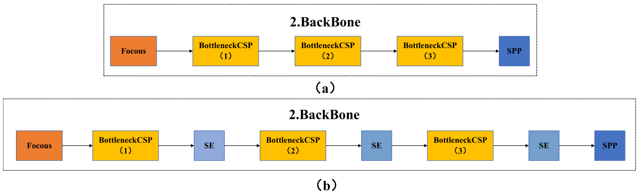


FIGURE 3. The backbone network of YOLOv5 integrates the SENet module.

activation layers. z_c is successively processed through dimension reduction and elevation by these two fully connected layers. The elevated output s retains the same shape as z_c, serving as the updated weight coefficients for the channels. The computation process is detailed in the subsequent equation:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma\left(W_2 \delta\left(W_1 z\right)\right). \tag{4}$$

After the first fully connected layer, a Rectified Linear Unit (ReLU) activation function is applied. In contrast, the subsequent fully connected layer utilizes a Sigmoid activation function. During training, iterative computations within the SENet module enable the two fully connected layers to establish a nonlinear relationship, thereby capturing the variances in the expressive capacities across channels.

The SENet module, based on the computed channel weight coefficients s, performs a reassignment operation on the input feature map U, enhancing the activation of relatively important channels. The process for this reassignment computation is detailed in the subsequent equation:

$$\tilde{x}_c = F_{\text{scale}}\left(u_c, s_c\right) = s_c \cdot u_c. \tag{5}$$

In the advanced layers of the YOLOv5 network, particularly within the Neck and Head segments, the feature maps carry more pronounced semantic information compared to the BackBone. Nonetheless, the SENet encounters difficulties distinguishing pivotal features from those small-scale feature maps that are characterized by a high degree of feature information fusion. This complexity hinders the apt allocation of learning weights. Even though the Backbone of YOLOv5 may not contain rich semantic information, it encapsulates essential texture and contour details of the targets. Such nuances are vital for detecting wind turbine blades which are devoid of distinct texture markers. By amplifying the learning weights associated with these texture and contour details, the efficacy of the model's training can be substantially augmented.

In our research, the SE module is amalgamated into the BackBone of the primary network, as delineated in Figure 3. Figure 3 (a) showcases the intrinsic main network structure of YOLOv5, while Figure 3 (b) portrays the configuration of the YOLOv5 main network post the SENet integration. The enhanced YOLOv5 algorithm with the integrated SENet is denominated YOLOv5-SENet. This iteration, despite a notable reduction in parameters, conserves the overarching information and bolsters the model's resilience.

SENet is a mechanism that discriminatively amplifies salient features while attenuating non-pertinent ones by harnessing global contextual information. This architecture facilitates the model in ascertaining feature weights contingent upon the incurred loss, thereby adjudicating the prominence of each feature map. In accordance with this discerned importance, a weight coefficient is ascribed to each feature channel. Such a stratagem enables the neural network to predominantly concentrate on the feature maps corresponding to segments of the wind turbine blade, augmenting their weight in the process. Concurrently, it deweights the less consequential or marginally influential feature maps, mitigating the perturbation from intricate backgrounds during model training and, thereby, enhancing the model's efficacy.

2) IMPROVEMENT OF LOSS FUNCTION

The loss function of YOLOv5 primarily consists of object loss, classification loss, and bounding box regression loss. Importantly, both the object loss and classification loss utilize BCE With Logits as their respective loss functions, as detailed below:

$$BCE(p, y) = \begin{cases} -\ln(p), & y = 1 \\ -\ln(1-p), & y = 0. \end{cases} \tag{6}$$

In the aforementioned equation, p represents the probability output after undergoing the Sigmoid activation function; y is the genuine sample label, taking values of either 0 or 1.

Within the image, regions containing the wind turbine are classified as positive samples, whereas the other areas are considered negative samples. For positive samples, a larger output probability corresponds to a reduced loss. In contrast, for negative samples, a smaller output probability results in a lesser loss. The imbalance between positive and negative samples is evident in one-stage object detection algorithms. Specifically, in drone-captured images of wind turbines, the background's proportion considerably exceeds that of the wind turbines. Consequently, the loss values produced by the loss function predominantly originate from the negative sample backgrounds. To address this, we have modified the loss function in this study, incorporating parameters to equilibrate the influence of both positive and negative samples on the loss.

To regulate the weights assigned to positive and negative samples, it's imperative to mitigate the influence of an abundance of negative samples on the loss. This objective can be accomplished using a balancing factor, as illustrated below:

$$\alpha_t = \begin{cases} \alpha, & y = 1 \\ 1-\alpha, & y = 0. \end{cases} \tag{7}$$

The factor a_t provides different weights depending on the sample label, and its principle is shown in the equation below:

$$BCE\left(p, y, \alpha_t\right) = \begin{cases} -\alpha \ln(p), & y = 1 \\ -(1-\alpha)\ln(1-p), & y = 0. \end{cases} \tag{8}$$

The factor a_t controls the proportion of positive and negative samples in the loss by adjusting its magnitude. When a_t is in the interval [0.50,1], it can increase the proportion of the positive sample loss while reducing that of the negative sample loss. An a_t value within the range [0.25,0.75] can achieve better AP (Average Precision) values.

The factor a_t is designed to regulate the contributions of positive and negative samples to the loss, without altering the loss pertaining to easily separable and challenging samples. Thus, modulation factors $(1-p)^{\gamma}$ and p^{γ} are employed to control the weights of the difficult-to-distinguish samples and easily distinguishable samples, with the underlying principle illustrated as follows:

$$BCE(p, y, r) = \begin{cases} -(1-p)^{\gamma}\ln(p), & y = 1 \\ -p^{\gamma}\ln(1-p), & y = 0. \end{cases} \tag{9}$$

In the aforementioned formula, γ falls within the range [0,5]. By adjusting the value of γ, one can control the magnitude of the modulation factor, thereby regulating the loss weight for hard-to-distinguish and easily distinguishable samples. When $\gamma = 0$, it represents the standard binary cross-entropy loss function. When $0 < \gamma \leq 5$, the effect is to reduce the contribution of easily classified samples to the loss, enabling the model to focus more on challenging samples.

By integrating the balancing factor a_t with the modulation factors $(1-p)^{\gamma}$ and p^{γ}, we obtain the improved Focal loss. The updated loss function is presented as follows:

$$\text{FocalLoss }(p, y, \alpha_t, \gamma) = \begin{cases} -\alpha(1-p)^{\gamma}\log(p), & y = 1 \\ -(1-\alpha)p^{\gamma}\log(1-p), & y = 0. \end{cases} \tag{10}$$

In this context, the balancing factor a_t addresses the imbalance between positive and negative samples in the model, thereby reducing the influence of complex backgrounds on model training. The modulation factors $(1-p)^{\gamma}$ and p^{γ} control the impact of the differences between easily and hard-to-distinguish samples on the loss. This enhances the discriminative power between segments of wind turbine blades, leading to improved training results for the model.

C. THE PRINCIPLE AND IMPROVEMENT OF DEEPSORT ALGORITHM

This study employs the DeepSORT algorithm [20], [21] for tracking wind turbine blades and assigning IDs, combined with the enhanced YOLOv5-SE, to distinguish the segments of the blades, aiming to identify specific segments of particular blades. Compared to its predecessor, SORT, DeepSORT significantly reduces ID loss by introducing cascade matching and new trajectory verification. This is crucial for this study, as it requires distinguishing between three blades that have extremely similar visual appearances. Large structures like wind turbines often exceed the camera's FOV and exhibit unique motion characteristics. Throughout the observation process, due to the movement of the blades, the region of interest frequently enters and exits the FOV, challenging the continuous and stable tracking of the wind turbine blades. This underscores the need to refine the Kalman filter, providing a robust solution for such complex scenarios.

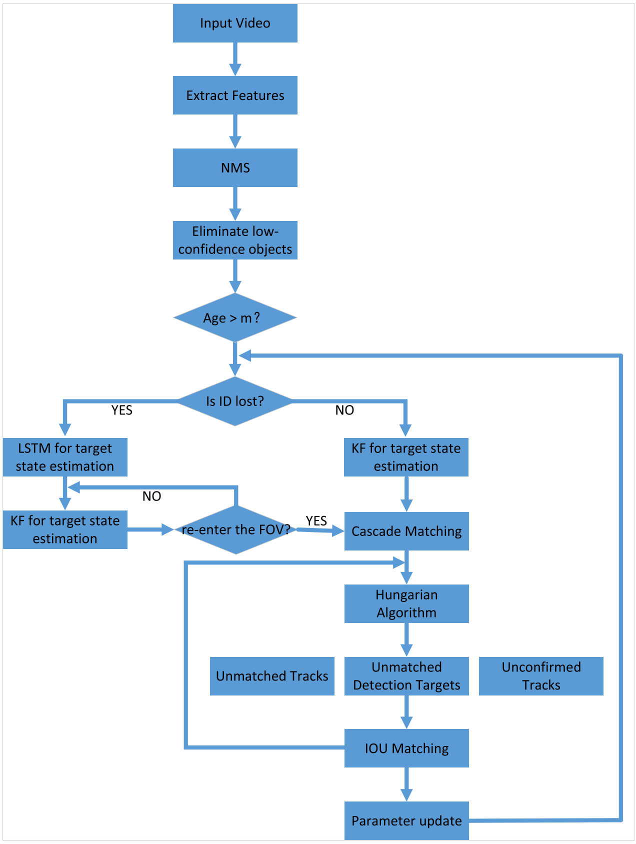


FIGURE 4. Optimized tracking flowchart.

In this study, YOLOv5 is employed as the detector and DeepSORT as the tracker. Further refinements were made to the Kalman filter within DeepSORT, integrating it with LSTM to enhance tracking robustness. The optimized tracking process is illustrated in Figure 4.

The entire tracking workflow is as follows:

(1) Object Detection: YOLOv5 is employed to detect the wind turbine blades in the input real-time video. For each video frame, YOLOv5 generates bounding boxes along with their respective confidence scores for all detected objects. Based on its confidence score, non-maximum suppression (NMS) is applied, eliminating detections with low confidence (below 0.6) and retaining high-confidence bounding boxes.

(2) LSTM Trajectory Prediction: If a target consecutively appears for more than a predefined threshold, denoted as 'm', an LSTM model is initialized to utilize the previously collected trajectory data for predicting the future trajectory of the target. If the target becomes occluded or exits the FOV, the LSTM model persists in generating trajectory predictions.

(3) Kalman Filter and Virtual Observation: Upon the target's departure from the FOV, the LSTM's trajectory

predictions serve as surrogate observations and are integrated into the Kalman filter. Using these surrogate observations, the Kalman filter continues estimating the state of the target until its re-entry into the FOV.

(4) Target Tracking and State Estimation: Within the FOV, the target's state and trajectory are estimated and predicted via the Kalman filter. For accurate target matching and tracking, the system leverages cascade matching combined with the Hungarian algorithm, ensuring that each target retains a unique identification ID.

(5) Identity Management and Update: The parameters of both the Kalman filter and LSTM model, alongside the state and identity information of the targets, undergo updates. As the target makes its re-entry into the FOV, the previously stored state and identity data facilitate prompt recognition and re-matching.

DeepSORT, by leveraging cascade matching and new trajectory confirmation techniques, has notably reduced instances of ID loss. This is especially critical for objects like wind turbine blades that exhibit unique motion characteristics. Its enhanced features play a pivotal role in maintaining tracking consistency and accuracy. When the blades exit the FOV, we introduce an LSTM model. This model utilizes prior trajectory data to generate potential future paths of the blades, addressing situations of target loss. These generated trajectories, considered as virtual observations, are fed into the Kalman filter. This ensures that the system can continue estimating the target state during periods of target loss until the blades re-enter the FOV. The integration and optimization of the Kalman filter offer a robust solution for this system. It is capable of adeptly handling the unique motion characteristics of wind turbine blades, ensuring system stability and tracking accuracy under diverse complex conditions. Finally, when the blades re-enter the FOV, the system can swiftly identify and rematch the target using previously saved state and identity information. This further guarantees persistent consistency in a specific region throughout the continuous tracking process.

By employing this technical approach that combines LSTM with DeepSORT, we have not only optimized the tracking process of wind turbine blades but also achieved sustained consistency within specific areas. This holds immense practical value for the stable operation and maintenance of wind turbines.

1) KALMAN FILTERING AND ITS ENHANCEMENTS

The Kalman filter [22], [23] is a linear recursive estimation method, widely used in estimating system states within time series data. Its fundamental assumptions are that system noise follows a Gaussian distribution and the system itself is linear or approximately linear. However, the rotational characteristics of wind turbine blades deviate from these assumptions.

The rotation of the turbine blades exhibits periodic properties, but the shape changes induced by rotation within images can compromise the continuous stability of tracking. Especially when the observational viewpoint focuses on specific parts of the blades, the frequent entrance and exit of the wind turbine blades from the FOV make tracking increasingly challenging.

Taking into account the aforementioned factors, this study designed an optimized Kalman filter model specifically for wind turbine blades. Based on the blade's angle and angular velocity, the research establishes the state vector as illustrated below:

$$X = \begin{bmatrix} \theta \\ \omega \end{bmatrix}. \tag{11}$$

This provides a foundation for the precise estimation of blade rotation. Furthermore, considering the nonlinear rotation characteristics of wind turbine blades, we optimized the prediction model. This optimization enables more accurate blade position predictions, thereby assisting DeepSORT in achieving more precise tracking.

From the perspective of a drone, the movement trajectory of the wind turbine blade can be considered as a type of perspective projection. For each detected blade, we can pinpoint the center of its bounding box. Due to perspective distortion, these centers describe a circular motion trajectory in the real world but are projected as an ellipse in the image. Observing this ellipse, we can estimate the distance from the bounding box's center to the turbine tower, deriving the major and minor axes of the ellipse. Crucially, this major axis actually represents the radius of the blade's circular motion in the real-world environment.

To ensure the accuracy of observations, our aerial shooting strategy aims to keep the drone and the wind turbine tower on the same horizontal plane as much as possible. In this manner, the wind turbine blade will mostly present as a vertical ellipse from the drone's viewpoint. In the mapping process from the ellipse to a circle, we can achieve a "stretching" correction for the perspective distortion by multiplying the x-coordinate by a factor k. For a point P on the ellipse, its mapped position on the circle is P'. The transformation process of its coordinates is shown as follows:

$$P' = \left(x', y'\right) = (kx, y). \tag{12}$$

wherein, k is the stretching factor, which is related to the camera's focal length, altitude, and other perspective parameters.

Using the position of the tower shaft as the origin O, we can compute the angle between point P' and O. Assuming the coordinates of O serve as the origin, the angle of the wind turbine blade can be obtained through the following equation:

$$\theta = \arctan\left(\frac{y'}{x'}\right). \tag{13}$$

The angular velocity is the rate of change of angle with respect to time. For two consecutive frames, the angular velocity can be obtained through the following equation:

$$\omega = \frac{\theta_{t+1} - \theta_t}{\Delta t}. \tag{14}$$

In the aforementioned equation, θ_{t+1} and θ_t are the angles in two consecutive frames, while Δt represents the time difference between the two frames.

Based on the aforementioned assumptions, the state transition model can be described as the following equation:

$$\mathrm{X}_{k+1} = \mathrm{AX}_k + \mathrm{BU}_k + \mathrm{W}_k. \tag{15}$$

In the aforementioned state transition model, A is the state transition matrix, U_k is the optional control input. However, for wind turbines, their motion is driven by wind speed, so we can typically disregard this input. W_k represents the process noise, which includes factors such as camera jitter and inherent blade vibrations.

For the prediction phase of the Kalman filter, we forecast based on the state from the previous moment. Given the non-linear rotational characteristics of the wind turbine blades, it's necessary to employ the Jacobian matrix for linearization to ensure accurate prediction. The uncertainty in the prediction is represented by the covariance P_k, and its covariance equation is shown as follows:

$$P_{k+1|k} = AP_kA^T + Q. \tag{16}$$

In the aforementioned equation, Q represents the covariance matrix of the process noise.

In the update phase, the Kalman gain K_k is used to determine the extent to which we should trust the prediction and the extent to which we should trust the observation. The formula for calculating the Kalman gain is as follows:

$$K_k = P_{k+1|k}H^T \left(HP_{k+1|k}H^T + R\right)^{-1}. \tag{17}$$

Finally, combining the prediction and observation, the updated state estimate and covariance can be obtained using the following equations:

$$X_{k+1} = X_{k+1|k} + K_k \left(Y_k - HX_{k+1|k}\right). \tag{18}$$

$$P_{k+1} = (I - K_kH)\, P_{k+1|k}. \tag{19}$$

Through this prediction-update loop, the optimized Kalman filter can provide effective estimation and tracking for nonlinear systems, especially in scenarios with noise and uncertainty. For tracking the wind turbine blades, this means that DeepSORT can consistently track their position and velocity, and quickly make corrections even when the blades move in and out of the FOV.

In summary, by improving the Kalman filter within DeepSORT, particularly for the specific scenario of wind turbine blades, we can significantly enhance tracking stability and accuracy, achieving consistent target calibration of key components of the blade under the motion state of the wind turbine.

2) UTILIZING LSTM IN CONJUNCTION WITH KALMAN FILTER

After the aforementioned optimization, the Kalman filter can now provide effective estimation and tracking. However, during the tracking process of wind turbine blades, tracking failures or losses frequently occur due to various factors such as occlusions and exiting the FOV. When the wind turbine blade leaves the FOV, its estimated trajectory is shown in Figure 5.

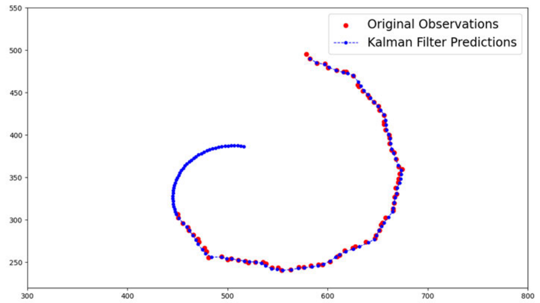


FIGURE 5. Kalman filter trajectory estimation diagram.

Figure 5 clearly illustrates that when trajectory estimation is performed with observation points, the optimized Kalman filter's trajectory estimation closely aligns with the distribution of the observation points. However, when the wind turbine blades leave the FOV, the lack of sufficient observational data to correct the prediction often results in the Kalman filter producing significant biases and uncertainties. This leads to a substantial deviation between the estimated trajectory and the actual trajectory. Such deviations can impact the DeepSORT algorithm's ability to correctly assign IDs when the blade re-enters the FOV.

To address the aforementioned issue, this study introduces an innovative approach that integrates the LSTM network [24] with the Kalman filter. The objective is to utilize the LSTM for accurate position estimation when the blade exits the FOV, ensuring the continuity of trajectory estimation by the Kalman filter.

When the blade is within the FOV, the Kalman filter operates normally, updating its state and covariance based on the dynamic behavior of the blade. Once the blade exits the FOV, the LSTM model intervenes, using previously collected trajectory data to generate a trajectory prediction for the period after the blade has left the FOV. Figure 6 below shows the training curve of the LSTM.

From Figure 6, it can be observed that the center coordinates of the turbine blade closely match the LSTM prediction model to the observed points. Once the blade leaves the FOV, the LSTM model will utilize the previously collected trajectory data to generate predictions for the blade's path outside of the FOV. These predicted data points serve as virtual observations, which are then fed into the Kalman filter to continuously update the blade's state. As a result, even in scenarios where the blade is lost from view, the combined LSTM and Kalman filter can offer a more stable and precise trajectory estimation, greatly enhancing the robustness of the DeepSort algorithm and laying a solid foundation for subsequent analysis and applications. The trajectory estimation using the Kalman filter in conjunction with LSTM is illustrated in Figure 7.

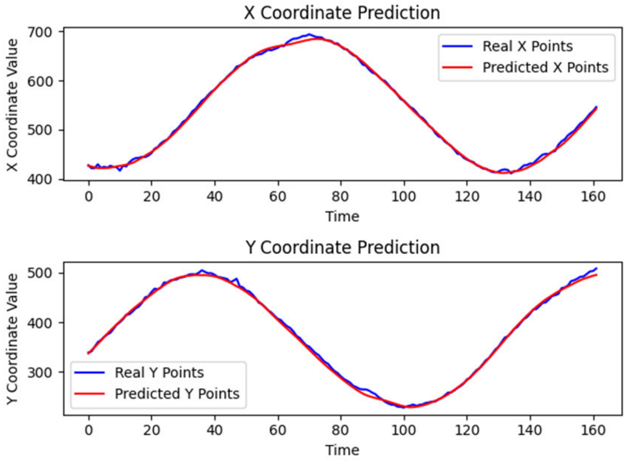


FIGURE 6. LSTM training curve diagram.

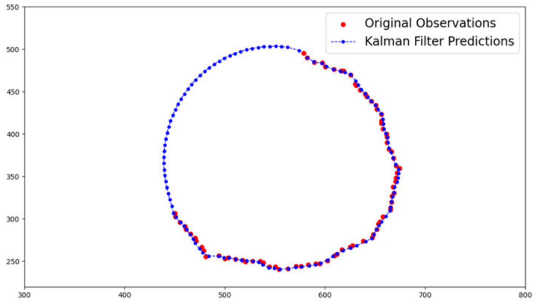


FIGURE 7. Trajectory estimation diagram with Kalman filter integrated with LSTM.

From Figure 7, it is evident that the trained LSTM model can predict the trajectory of the wind turbine blades. Due to its superior capability in handling sequential data, the model is able to learn the inherent patterns of blade motion, generating accurate trajectory predictions even when the blade exits the FOV. The Kalman filter, in turn, utilizes these virtual observations, achieving a seamless estimation of the blade's state.

Through this approach, the robustness of the DeepSort algorithm during blade occlusions is significantly enhanced, realizing a more coherent and accurate trajectory estimation, laying a solid foundation for subsequent analysis and applications. This method not only stabilizes the tracking algorithm but also ensures more precise and reliable localization when the blade re-enters the FOV, effectively overcoming the challenges posed by blade occlusions and exits from the visual field.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL ENVIRONMENT

The experimental environment used for pedestrian target tracking in this paper is Ubuntu18.04 operating system, Python 3.7.0, 8G RAM, Intel(R) CoreTMi9-10900K processor, and NVIDIA GeForce RTX 3090 (8G) GPU. The camera used in this experiment is Logitech StreamCam, with a maximum video resolution of 720p/30fps.

In this study, data collection was carried out in a wind farm located in an eastern coastal region of China. A total of 6089 video frames were captured. To avoid minimal changes between adjacent frames, 3451 frames were selected from these as the experimental dataset for this research. Two experimental scenarios were designed: one with a whole wind turbine sequence of 960 frames and a local sequence of 2491 frames. After completing the dataset construction and organization, the custom dataset was uploaded and made public. This dataset can be found on the public repository IEEE DataPort and can be accessed and downloaded through a link. The dataset was manually annotated using the LabelImg annotation tool, eventually generating labels in txt format, with coordinates all undergoing normalization. The tip, middle, and base of the wind turbine blades are represented by the numbers 0, 1, and 2, respectively. Images were divided into training, validation, and test sets at a ratio of 8:1:1.

B. ANALYSIS OF MODEL TRAINING AND DETECTION RESULTS

To test the performance of the fused model after adding SEnet, this study trained the network model using a dataset established for wind turbines, employing the same environment and parameter configurations.

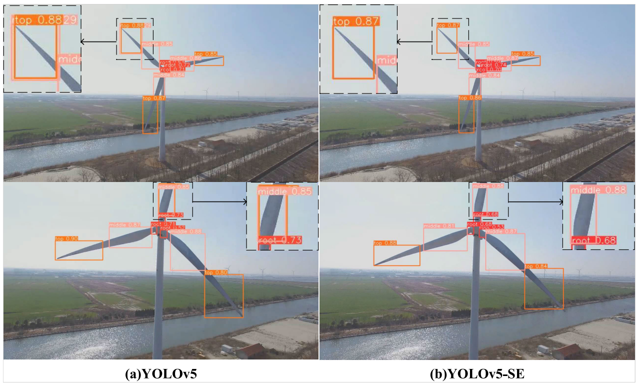
To more intuitively evaluate the performance of the proposed fused model in detecting segments of wind turbine blades — especially areas lacking texture features — we designed and carried out a series of visualization experiments. The results of these experiments are detailed in Figure 8.

Compared to the original YOLOv5 model, the fusion model developed in this study shows its superiority in multiple aspects. Specifically, the original model struggles with distinguishing between the tip and the middle sections of the wind turbine blades, often confusing these two parts. This issue is particularly pronounced when the blade's tip and middle sections overlap with other parts of the turbine due to their high similarity in appearance, which significantly complicates the recognition process.

However, the fusion model proposed in this study offers effective solutions to these challenges. Not only does the model accurately identify different parts of the wind turbine blades, but it also exhibits strong robustness against complex backgrounds and overlapping areas.

To comprehensively assess the model's performance, this study adopted various evaluation metrics, including mAP_0.5, Precision, and Recall. Furthermore, a series of visualization steps were conducted to provide a more intuitive display of the model's performance trends. Figure 9 illustrates the change in these evaluation metrics as the number of training iterations increases.

From the data in Figure 9, it is evident that, compared to the original YOLOv5 model, the integrated model demonstrates a faster convergence rate across all evaluation metrics. Notably, this performance enhancement is achieved with a reduction in the model's feature parameters, further attesting to the



(a)YOLOv5 (b)YOLOv5-SE

FIGURE 8. Algorithm detection and recognition results.

TABLE 1. Detection model performance metrics.

Model	top	middle	root	mAP_0.5	Precision	Recall	FPS
YOLOv5	0.99	0.944	0.807	0.914	0.916	0.912	153
YOLOv5-SE	0.995	0.995	0.91	0.967	0.978	0.980	143

efficiency of the optimization. The integrated model exhibits a significant improvement in accuracy and robustness in the local detection of wind turbine blades. These experimental results not only validate the high accuracy of the fused model in handling highly similar or overlapping objects but also lay a solid technical foundation for its broader adaptability in practical applications.

To systematically assess the effectiveness of the improved YOLOv5 backbone network proposed in this study, we conducted a series of comparative experiments and consolidated the results in Table 1.

After a detailed analysis of the data in Table 1, we draw the following conclusions: The integrated model with SEnet achieved a significant improvement in the mAP_0.5. Specifically, its mAP_0.5 reached 96.7%, a 5.3% increase compared to the original YOLOv5 model's 91.4%. This result clearly indicates that the introduction of SE-net significantly enhanced the performance of the feature extraction network, enabling the model to make more efficient and rational inferences in complex environments. Notably, despite the addition of extra network structures, the fused model's detection speed still remains at 143 frames per second. This not only demonstrates the model's high computational efficiency but also meets the requirements for real-time applications.

C. TRACKING RESULTS AND ANALYSIS

To comprehensively evaluate the effectiveness of enhancing the DeepSORT tracking algorithm by combining LSTM for predicting blade center trajectory and optimizing the Kalman filter, this study designed a series of comparative experiments, particularly focusing on the dynamic tracking of wind turbine blades. The performance of DeepSORT is typically assessed by two main metrics: MOTP (Multiple Object Tracking Precision) and MOTA (Multiple Object Tracking Accuracy). MOTP reflects the positional precision of the correctly tracked targets, measuring the average error between the positions of the tracked objects and their actual positions. MOTA takes into account multiple factors to evaluate the accuracy of tracking, including false positives (FP), false negatives (FN), and identity switches (IDS). The formulas for calculating MOTP and MOTA are as follows:

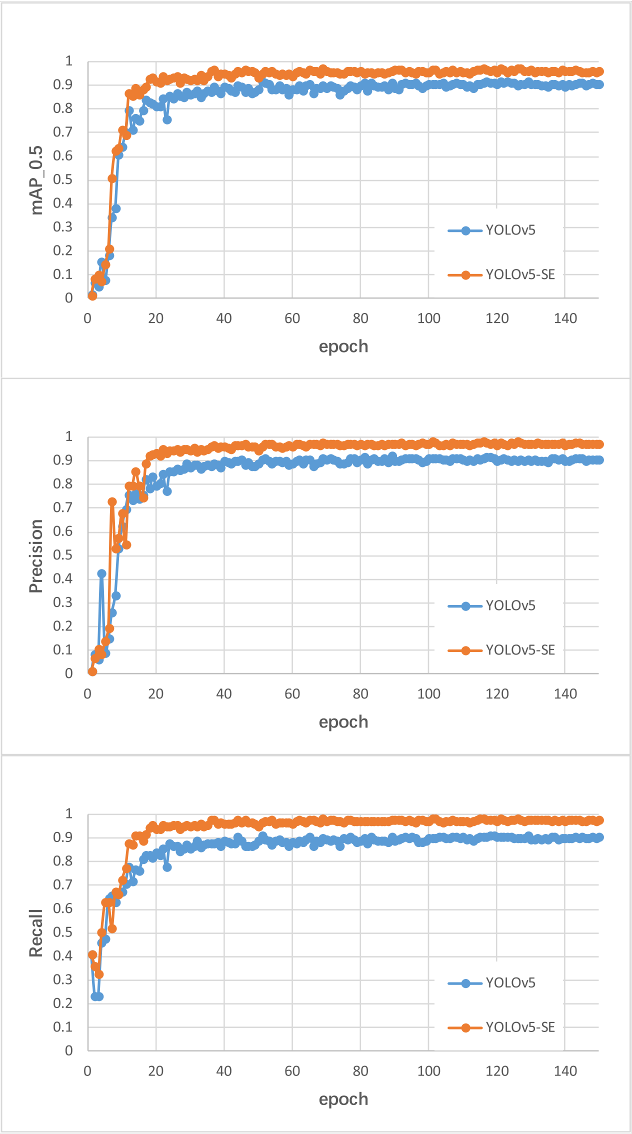
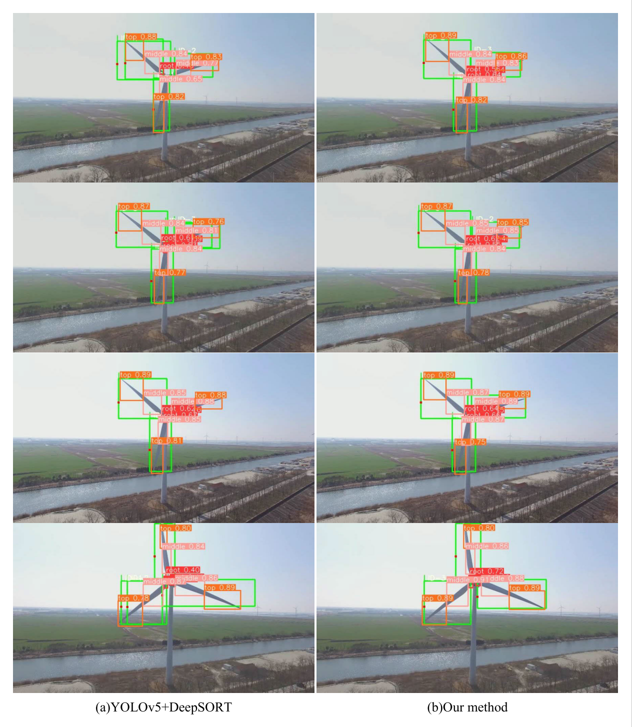


FIGURE 9. Variation curve of model metrics.

$$MOTP = \frac{\sum_{i,t} d_{i,t}}{\sum_t TP_t}. \tag{20}$$

$$MOTA = 1 - \frac{\sum_t (FP_t + FN_t + IDS_t)}{\sum_t GT_t}. \tag{21}$$

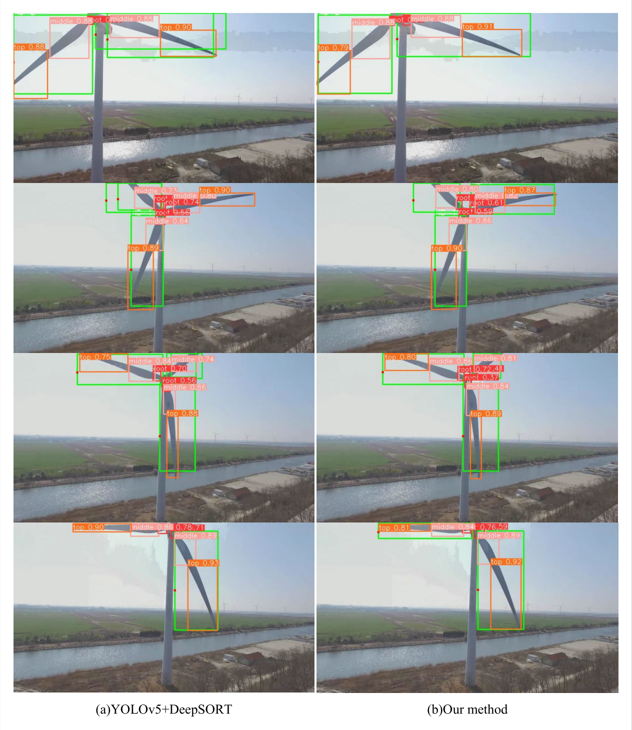
Scenario one primarily targets the global tracking of wind turbine generators. The experiment ensures the complete visibility of the wind turbine generator within the camera's FOV by hovering the drone in front of the wind turbine tower. As shown in Figure 10, it presents the integrated performance of the two tracking algorithms in a real-world wind turbine



(a)YOLOv5+DeepSORT

(b)Our method

FIGURE 10. Global tracking results of the wind turbine.



(a)YOLOv5+DeepSORT

(b)Our method

FIGURE 11. Local tracking results of the wind turbine.

scenario. Given that the wind turbine generator rotates clockwise, the ID values are arranged counterclockwise from smallest to largest, providing an intuitive description of the motion. The colorful detection boxes (in orange, pink, and red) not only differentiate the different sections of the blade (top, middle, and root) but also provide category information and detection accuracy.

Figure 10 shows a comparative demonstration where the original DeepSORT algorithm exhibits limitations when tracking wind turbines, such as imprecise localization of the tracking box and the presence of redundant boxes. These issues could potentially lead to subsequent misidentification and reduced robustness. After integrating the movement information of the wind turbine, the performance of the DeepSORT algorithm is noticeably enhanced, especially in the localization of the tracking box and reduction of misidentification, thereby augmenting the algorithm's accuracy and robustness.

In scenario two, a drone hovers at a close distance in front of the wind turbine, causing the camera's FOV to mainly focus on the local region of the wind turbine, where the blades frequently enter and exit the viewpoint. This design aims to validate the tracking performance of the optimized Kalman filter in the DeepSORT algorithm, particularly when blades frequently enter and exit the visual field.

The results from Figure 11 clearly highlight the evident shortcomings of the original tracking model when the turbine blades frequently enter and exit the visual field. When the blades leave and subsequently re-enter the FOV, the model often loses its tracking frame, resulting in inconsistencies in object identification. In contrast, the DeepSORT tracking model improved with LSTM integration performs outstandingly in such scenarios of high-frequency entry and exit. Not only does it accurately label the turbine blades, but it also tracks their trajectory continuously, significantly reducing the risk of target loss.

TABLE 2. Tracking model performance metrics.

Model	Scenarios	FP	FN	IDS	MOTP%	MOTA%
DeepSORT	Global	241	743	41	77.7	64.4
	Local	387	1141	62	77.5	61.7
Our method	Global	198	701	34	78.1	67.6
	Local	292	1072	43	77.9	66.1

This substantial advantage can be primarily attributed to the refinement of the Kalman filter integrated with LSTM. Traditional Kalman filters predict the next state of an object based on linear assumptions, but such predictions can be inaccurate under certain complex motion patterns. LSTM, being a long short-term memory network, can better learn and predict complex patterns in time series. In this application, the integration of LSTM allows the Kalman filter to predict the turbine blade's movement trajectory with higher accuracy, thereby enhancing the tracking algorithm's precision and robustness.

Table 2 further contrasts the tracking performance of the blades using the original DeepSORT algorithm and our method under two distinct scenarios.

The data from Table 2 clearly demonstrates that the LSTM-enhanced version of DeepSORT outperforms its original counterpart, especially in scenarios with frequent entries and exits from the FOV, exhibiting greater robustness in reducing misidentifications and target losses.

Overall, the experimental results validate the superiority of the proposed dynamic target tracking algorithm at a superfield scale, integrating deep learning with global motion information, in tracking the blades of wind turbines.

V. CONCLUSION

Through the improved YOLOv5 object detection algorithm and the DeepSORT object tracking algorithm, this study successfully achieved consistent target labeling of the blades of wind turbines in operation and their specific parts. Given the high similarity among the blades and the lack of distinct texture features between the blade segments, we incorporated the SEnet into the backbone network of YOLOv5 and optimized the focal loss function. Experimental results indicate that the enhanced model has achieved a 5.3% improvement in the mAP evaluation metric. During the tracking process, the frequent entry and exit of the blades from the drone's FOV could jeopardize tracking stability. To address this issue, we optimized the Kalman filter in DeepSORT based on the motion patterns and characteristics of the wind turbines and incorporated LSTM to predict the position of the blade once it exits the FOV. This improvement allows the system to track specific blades continuously and stably, enhancing tracking robustness even when blades frequently enter and exit the visual field. In summary, the method proposed in this study has successfully achieved consistent target labeling of crucial components of specific turbine blades under motion, offering technical support for the widespread application of drones in wind turbine inspections. Although our approach demonstrates exceptional performance in the two outlined scenarios, there are inherent limitations. For instance, should there be an abrupt change in the motion of the wind turbine assembly, the LSTM might necessitate a certain period to adapt to such a shift. In the future, we aim to incorporate more sophisticated deep learning models, like the Transformer, to further bolster the robustness of tracking.

REFERENCES

[1] D. Xu, C. Wen, and J. Liu, "Wind turbine blade surface inspection based on deep learning and UAV-taken images," *J. Renew. Sustain. Energy*, vol. 11, no. 5, Sep. 2019, Art. no. 053305.

[2] B. Yang and D. Sun, "Testing, inspecting and monitoring technologies for wind turbine blades: A survey," *Renew. Sustain. Energy Rev.*, vol. 22, pp. 515–526, Jun. 2013.

[3] J. S. Hwang, D. J. Platenkamp, and R. P. Beukema, "A literature survey on remote inspection of offshore wind turbine blades: Automated inspection and repair of turbine blades (AIRTuB)-WP1," Roy. Netherlands Aerosp. Centre, Amsterdam, The Netherlands, Tech. Rep. NLR-CR-2020-223-RevEd-1, May 2021.

[4] M. Stokkeland, K. Klausen, and T. A. Johansen, "Autonomous visual navigation of unmanned aerial vehicle for wind turbine inspection," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, Jun. 2015, pp. 998–1007.

[5] S. Lu, B. Wang, H. Wang, L. Chen, M. Linjian, and X. Zhang, "A real-time object detection algorithm for video," *Comput. Electr. Eng.*, vol. 77, pp. 398–408, Jul. 2019.

[6] X. Ran, S. Zhang, H. Wang, and Z. Zhang, "An improved algorithm for wind turbine blade defect detection," *IEEE Access*, vol. 10, pp. 122171–122181, 2022.

[7] Z. Xiaoxun, H. Xinyu, G. Xiaoxia, Y. Xing, X. Zixu, W. Yu, and L. Huaxin, "Research on crack detection method of wind turbine blade based on a deep learning method," *Appl. Energy*, vol. 328, Dec. 2022, Art. no. 120241.

[8] X. Yang, Y. Gao, C. Fang, Y. Zheng, and W. Wang, "Deep learning-based bolt loosening detection for wind turbine towers," *Struct. Control Health Monitor.*, vol. 29, no. 6, Jun. 2022, Art. no. e2943.

[9] R. Zhang and C. Wen, "SOD-YOLO: A small target defect detection algorithm for wind turbine blades based on improved YOLOv5," *Adv. Theory Simul.*, vol. 5, no. 7, Jul. 2022, Art. no. 2100631.

[10] L. Yang, J. Fan, S. Song, and Y. Liu, "A light defect detection algorithm of power insulators from aerial images for power inspection," *Neural Comput. Appl.*, vol. 34, no. 20, pp. 17951–17961, Oct. 2022.

[11] K. Xia, Z. Lv, C. Zhou, G. Gu, Z. Zhao, K. Liu, and Z. Li, "Mixed receptive fields augmented YOLO with multi-path spatial pyramid pooling for steel surface defect detection," *Sensors*, vol. 23, no. 11, p. 5114, May 2023.

[12] Y. Wang, C. Tang, J. Wang, Y. Sang, and J. Lv, "Cataract detection based on ocular B-ultrasound images by collaborative monitoring deep learning," *Knowl.-Based Syst.*, vol. 231, Nov. 2021, Art. no. 107442.

[13] S. Chen, J. Lan, H. Liu, C. Chen, and X. Wang, "Helmet wearing detection of motorcycle drivers using deep learning network with residual transformer-spatial attention," *Drones*, vol. 6, no. 12, p. 415, Dec. 2022.

[14] J. Kim, "Obstacle information aided target tracking algorithms for angle-only tracking of a highly maneuverable target in three dimensions," *IET Radar, Sonar Navigat.*, vol. 13, no. 7, pp. 1074–1080, Jul. 2019.

[15] Y. Ning, L. Zhao, C. Zhang, and Z. Yuan, "STD-YOLOv5: A ship-type detection model based on improved YOLOv5," *Ships Offshore Struct.*, pp. 1–10, Nov. 2022.

[16] Y. J. Wang and G. Huang, "Target tracking algorithm of basketball video based on improved grey neural network," *Sci. Program.*, vol. 2021, pp. 1–8, Oct. 2021.

[17] J. Rao, K. Xu, J. Chen, J. Lei, Z. Zhang, Q. Zhang, W. Giernacki, and M. Liu, "Sea-surface target visual tracking with a multi-camera cooperation approach," *Sensors*, vol. 22, no. 2, p. 693, Jan. 2022.

[18] D. Du, "Multiperson target dynamic tracking method for athlete training based on wireless body area network," *Adv. Math. Phys.*, vol. 2021, pp. 1–9, Nov. 2021.

[19] F. Jubayer, J. A. Soeb, A. N. Mojumder, M. K. Paul, P. Barua, S. Kayshar, S. S. Akter, M. Rahman, and A. Islam, "Detection of mold on the food surface using YOLOv5," *Current Res. Food Sci.*, vol. 4, pp. 724–728, Oct. 2021.

[20] S. Kapania, D. Saini, S. Goyal, N. Thakur, R. Jain, and P. Nagrath, "Multi object tracking with UAVs using deep SORT and YOLOv3 RetinaNet detection framework," in *Proc. 1st ACM Workshop Auto. Intell. Mobile Syst.*, Jan. 2020, pp. 1–6.

[21] T. L. Dang, G. T. Nguyen, and T. Cao, "Object tracking using improved Deep_SORT_YOLOv3 architecture," *ICIC Exp. Lett.*, vol. 14, no. 10, pp. 961–969, 2020.

[22] G. F. Welch, "Kalman filter," in *Computer Vision: A Reference Guide*. New York, NY, USA: Springer,2020, pp. 1–3.

[23] G. Welch and G. Bishop, "An introduction to the Kalman filter," Dept. Comput. Sci., Univ. North Carolina, Chapel Hill, NC, USA, Tech. Rep. TR 95-041, 1995.

[24] N. K. Manaswi and N. K. Manaswi, "RNN and LSTM," in *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras*. Berkeley, CA, USA: Apress, pp. 115–126, 2018.

YIMING XU received the B.S., M.S., and Ph.D. degrees from the Nanjing University of Science and Technology, China, in 2003, 2005, and 2011, respectively. Since 2011, he has been with Nantong University. He became a Professor, in 2022. His research interests include artificial intelligence and advanced sensing technology.

ZHENYU FU is currently pursuing the Master of Engineering degree with Nantong University, primarily focused on digital image processing and its applications in computer vision. His work concentrates on leveraging UAV technology for the inspection of wind turbine assemblies, providing substantial technical support for the widespread application of this technology in the field.

WEI PENG received the Master of Engineering degree from Nantong University. He has delved into the intricacies of UAV-based digital image processing and its applications with Nantong University. His research is dedicated to exploring how unmanned aerial vehicle technology can enhance the quality and efficiency of image processing. His aim is to propel the innovative application of UAV technology in critical areas, such as intelligent surveillance, geographic information system mapping, agricultural inspection, and disaster assessment. His work not only demonstrates the immense potential of UAV technology in image acquisition and analysis but also provides valuable scientific data for the future development and optimization of UAV applications.

ZIHENG DING is currently pursuing the Master of Engineering degree with Nantong University, specializing in the in-depth study of object detection and tracking through deep learning methods. His research interests include developing advanced algorithms to improve the precision and efficiency of computer vision systems.

GUAN LU received the B.S. degree from Southeast University, China, in 2005, and the M.S. and Ph.D. degrees from the Nanjing University of Aeronautics and Astronautics, China, in 2008 and 2011, respectively. Since 2011, she has been with Nantong University. She became a Professor, in 2022. Her research interests include intelligent manufacturing and measurement and control technology.

QIANG LIU received the Ph.D. degree in computer science from the University of Essex, U.K. He is currently a Lecturer with the Department of Engineering Mathematics and Technology, University of Bristol. Additionally, he holds the position of a Honorary Research Scientist with the Department of Psychiatry, University of Oxford, where he conducted his postdoctoral research. His research interests include machine learning algorithms, especially deep neural networks, sensor integration and data fusion algorithms, and natural language processing.