IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 August 2024, accepted 8 August 2024, date of publication 13 August 2024, date of current version 22 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3442786

RESEARCH ARTICLE

YOLO-Based Missile Pose Estimation Under Uncalibrated Conditions

CHANGHONG JIANG[1], XIAOQIAO MU[2], BINGBING ZHANG[3], MUJUN XIE[1], AND CHAO LIANG[4]

[1]School of Electrical and Electronic Engineering, Changchun University of Technology, Changchun, Jilin 130012, China
[2]School of Mechanical and Electrical Engineering, Changchun University of Technology, Changchun 130012, China
[3]School of Computer Science and Engineering, Dalian Minzu University, Dalian 116602, China
[4]College of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China

Corresponding authors: Mujun Xie (xiemujun@ccut.edu.cn) and Chao Liang (liangchao@ccut.edu.cn)

This work was supported by the Science and Technology Development Program Project of Jilin Province under Grant 20230201111GX and Grant 20230201039GX.

ABSTRACT In missile docking, high-precision section alignment is vital for mission success. Traditional techniques, relying on radar and GPS, face calibration complexities and environmental interference risks, potentially leading to inaccurate estimations. To address this issue, the PoseNoCal-YOLOv5 is developed, offering a vision-based pose estimation approach that requires no traditional calibration, enhancing docking precision and reliability. It comprises two sub-networks: an improved YOLOv5 object detection model with an attention mechanism for precise object detection, and an uncalibrated pose estimation module using re-linearization for pose estimation without camera calibration. A simulated dataset is created for validation, covering diverse docking scenarios. Extensive experiments on this simulated datasets prove the effectiveness of the proposed method.

INDEX TERMS Missile docking, vision-based pose estimation, YOLOv5s, uncalibrated agorithm, attention mechanism.

I. INTRODUCTION

Mi ssile technology is a key component of the modern defense industry, and its level of development largely determines a country's strength in international competition. The quality and quantity of missiles have become important indicators of national power. The final assembly of missiles is a critical and core technology in the missile manufacturing process, and the docking of missile sections is the main content of the missile's final assembly. Traditional missile docking assembly often employs laser [1], [2], [3] and GPS technologies [4]. These methods have a low level of automation, are sensitive to environmental conditions, are costly, and involve complex equipment, which cannot meet the efficiency and economic requirements of modern missile production and manufacturing. Therefore, the development of efficient and reliable missile docking technology is of strategic significance. This paper adopts a vision-based pose estimation method [5], [6]. A novel algorithm called PoseNoCal-YOLOv5 is proposed in this paper, which has the advantages of high cost-effectiveness and simple equipment. The PoseNoCal-YOLOv5 algorithm consists of two core components: an improved target detection [7] and an uncalibrated pose measurement algorithm [9], designed to provide a precise and efficient solution for missile section docking. To comprehensively verify the effectiveness of the PoseNoCal-YOLOv5 algorithm, we have constructed a simulated missile dataset.

The following are the three main innovative aspects of the algorithm:

- The PoseNoCal-YOLOv5 algorithm employs an enhanced object detection algorithm that, when combined with the SE module, improves the accuracy of target identification and localization. This integration not only enhances the algorithm's detection capabilities in complex environments but also, through the adaptive rescaling

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

mechanism of the SE module, increases the precision of identifying targets of various scales.

- A core innovation of the algorithm is the introduction of an uncalibrated pose estimation module that estimates the position and orientation of missile sections directly from images without the need for complex camera calibration. This innovation significantly reduces the complexity and cost of system deployment while ensuring the accuracy and real-time nature of pose estimation.

- To comprehensively validate the effectiveness of the PoseNoCal-YOLOv5 algorithm, we conducted accuracy verification on the widely recognized COCO128 public dataset. Furthermore, we constructed two specialized datasets for testing. The first dataset, generated through code simulation, contained error-free ideal data, serving as a benchmark for comparative analysis. The second dataset, generated through software simulation, encompassed a variety of scenarios that could be encountered during the actual missile docking process. The algorithm exhibited outstanding performance on these datasets, signifying a notable progression in the domain of missile automation and intelligent systems.

II. RELATED WORK

A. RESEARCH STATUS OF MISSILE DOCKING POSE ESTIMATION

Missile docking pose estimation technology plays a critical role in the missile docking process [39], primarily focused on acquiring the relative position and pose information of the missile's various components for precise docking control. Early pose estimation techniques primarily utilized mechanical sensors [2], [3], [4], [34], [35], [37] (e.g., encoders, gyroscopes) and optical sensors [5], [6] (e.g., laser rangefinders, optical theodolites). These techniques partially satisfy the accuracy requirements of missile docking, however, their principles and structures constrain their performance, leading to certain errors and drawbacks, such as susceptibility to environmental interference, complex installation procedures, and high costs.

B. DEEP LEARNING FOR OBJECT DETECTION

In the past decade, target detection technology has undergone significant development and progress. Since 1998, the rise of Convolutional Neural Networks(CNNs) has brought revolutionary changes to the field of image recognition and processing [10]. In 2012, the AlexNet model, based on CNNs, achieved breakthrough results in the ImageNet Large Scale Visual Recognition Challenge, marking the rise of deep learning in the field of image recognition [11]. Subsequently, in 2014, researchers from the University of Oxford and Google collaborated to introduce the VGGNet [12]. VGGNet demonstrated exceptional performance in image classification and object localization tasks by deepening the network structure, further confirming the potential of deep networks in enhancing model performance. However, as networks deepened, new challenges emerged, such as a decline in model accuracy. To overcome these challenges, researchers proposed a variety of innovative solutions. Among them, the ResNet [13] proposed by He et al. in 2015 successfully addressed the degradation problem in deep network training by introducing residual learning mechanisms. In the field of target detection, the R-CNN [14] model, as the pioneering work, achieved significant results in accuracy but was slow in detection speed. To address this issue, researchers developed the Fast R-CNN [15] and Faster R-CNN [16], which significantly improved the speed of target detection by optimizing network structures and detection processes. In 2016, the introduction of the YOLO (You Only Look Once) network [17] brought a new perspective to the field of target detection. YOLO achieved fast and accurate target detection by simplifying the detection process, eliminating the need for complex candidate region extraction steps. At the same time, numerous innovative algorithms based on YOLO have also been proposed. [36]Additionally, a series of innovative network models, such as Feature Pyramid Networks(FPN) [18], Mask R-CNN [19], Cascade R-CNN [20], and Fully Convolutional One-Stage Object Detection(FCOS) [21] have made improvements and innovations in different aspects of target detection, advancing the development of the technology.

In our analysis, YOLOv5 was compared against YOLOv4, SSD, and Faster R-CNN on the COCO 2017 dataset. The results showed that YOLOv5 achieves an $\mathbf{mAP_{50}}$ of 94.7% and an F1 score of 87.6%, outperforming other methods. Its inference speed of 45 FPS also surpasses the competition, highlighting its real-time capabilities. The choice of YOLOv5 as our base model was driven by its blend of speed and precision, essential for real-time applications.

C. PNP ALGORITHMS FOR EFFICIENT POSE ESTIMATION

In the domain of computer vision and robotics, the PnP (Perspective-n-Point) problem [22] is pivotal for ascertaining the camera's pose based on correspondences between 3D points and their 2D projections. Throughout the years, a variety of algorithms have emerged, each aiming to address this challenge with varying levels of efficacy and efficiency. In 2016, two foundational papers [23], [24] delivered a thorough examination of image-based camera localization techniques, covering methods applied to both known and unknown environments. These works set the stage for future research by proposing geometric solutions to the PnP problem. Subsequently, an innovative approach [25] surfaced, concentrating on rapid and robust absolute camera pose estimation when the camera's focal length was known in 2018. This technique capitalized on the known focal length to enhance the precision and velocity of pose estimation. In the same year, a further significant contribution [26] to the field presented a simplified geometric solution for the PnP problem, making the process more straightforward and practical. The year 2021 witnessed the emergence of three distinct papers [27], [28], [29] that

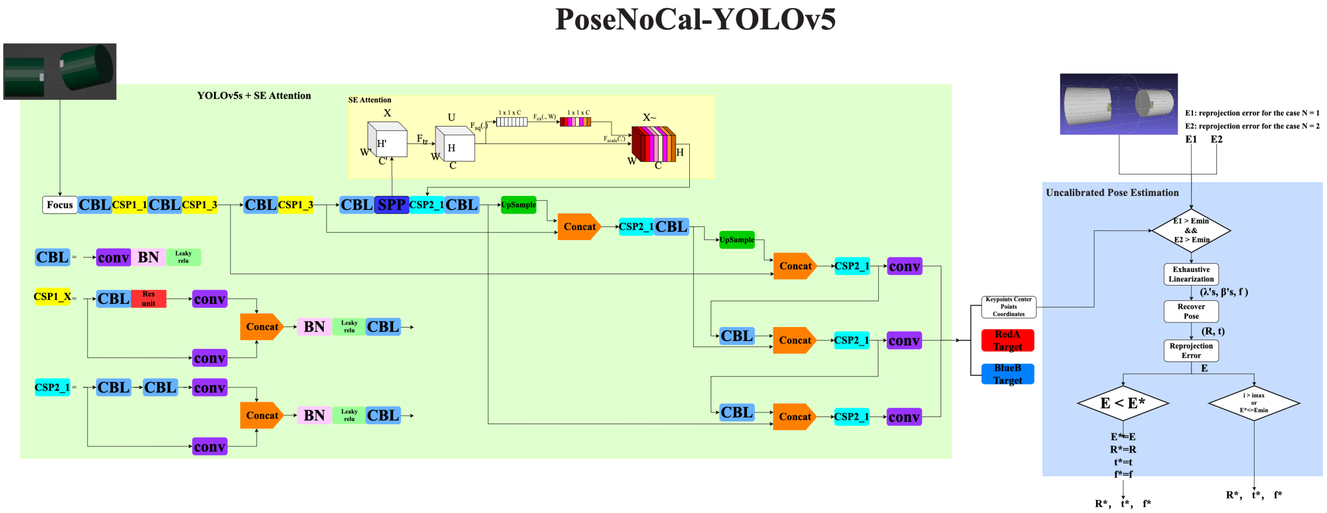

FIGURE 1. Algorithm overall.

pushed the boundaries of PnP problem-solving. One paper proposed an efficient DLT (Direct Linear Transform)–based method for tackling the PnP and its variant problems. Another introduced an uncertainty-aware camera pose estimation method that leveraged both point and line features. The third paper presented an accurate and efficient solution for the PnPL(Perspective-n-Line) problem through null space analysis. In 2022, a study [30] continued this progress by proposing an accurate and efficient solution for the PnPL problem, also employing null space analysis. This research refined the techniques for handling line correspondences in pose estimation.

III. METHOD

This study introduces an innovative pose estimation method called PoseNoCal-YOLOv5, designed to enhance the accuracy and efficiency of target localization. As shown in Figure 1, the algorithm consists of two key components: The first is an improved target detection algorithm based on YOLOv5s, which integrates an SE (Squeeze-and-Excitation) module to significantly enhance the recognition and extraction of image features; The second part is an uncalibrated pose estimation algorithm, a novel module that abandons the cumbersome camera calibration steps of traditional methods. It directly utilizes the keypoint information provided by the first part, combined with efficient computational techniques, to accurately calculate the pose of the target object. This innovative approach not only simplifies the calibration process but also greatly reduces the reliance on specialized knowledge and equipment, making pose estimation faster, more intuitive, and suitable for a variety of complex environments.

A. ENHANCED OBJECT DETECTION MODEL

The incorporation of the Squeeze-and-Excitation (SE) module into the YOLOv5s architecture enhances the model's capacity to discern and emphasize features critical for object detection. The enhanced object detection network is shown in Figure 1. Operating through a two-step process, the SE module initially condenses the spatial dimensions of feature maps using global average pooling, yielding a channel descriptor rich with global spatial insights. Subsequently, a series of fully connected layers, constituting the Excitation phase, are employed to capture and model the intricate interdependencies among channels. This process results in a set of weights that recalibrate the channel responses, amplifying salient features while suppressing irrelevant or redundant information. This recalibration is pivotal as it allows the network to reuse features more effectively, enriching the feature representation and enabling a more nuanced understanding of the input data. By dynamically adjusting to the significance of various features, the SE module endows the network with a flexibility that is crucial for accurate object detection. This dynamic adjustment not only bolsters the network's robustness against noise but also refines its focus on the most informative features, leading to a more precise detection outcome. While the SE module slightly increases the model's computational complexity, the strategic enhancement of feature relevance and the subsequent reduction in unnecessary computations can lead to a net improvement in the overall efficiency of the object detection process. In essence, the integration of the SE module into YOLOv5s is a theoretically sound approach to achieving a more accurate and efficient object detection model.

B. UNCALIBRATED POSE ESTIMATION ALGORITHM

This paper[9] presents an uncalibrated pose estimation algorithm. The uncalibrated pose and focus length estimation algorithm addresses the challenge of estimating a camera's pose (rotation matrix, translation vector, and focal length) without the need for calibration information, by utilizing a set of 3D-to-2D point correspondences. The estimation of the camera's pose is accomplished by analyzing the 3D-to-2D point correspondences. To estimate the camera pose and focal length from the detected corresponding key points, we initially constructed a linear system based on each set of correspondences as Equation (1). By analyzing the system matrix through Singular Value Decomposition (SVD), we determined the rank of the null space. When the rank is one, we can directly solve for the unknown vector. However, when the rank exceeds one, we employ exhaustive linearization and linearization methods, which are strategies for systematically exploring the solution space to find a solution that minimizes the reprojection error. This approach is not only applicable to data with noise but is also capable of handling a large number of correspondences, thereby enhancing the robustness and accuracy of the algorithm.

$$s \cdot \begin{bmatrix} u_i - u_0 \\ v_i - v_0 \end{bmatrix} = \begin{bmatrix} f \; 0 \; -\frac{x_i}{z_i} \\ 0 \; f \; -\frac{y_i}{z_i} \end{bmatrix} \cdot \begin{bmatrix} R_{13} \\ R_{23} \\ R_{33} \end{bmatrix} \cdot \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{1}$$

where s is the scaling factor used to handle the case where the focal length is unknown. (u_0, v_0) are the principal point coordinates of the image. f is the unknown focal length. (x_i, y_i, z_i) are the coordinates of a point in 3D space. In the camera pose estimation problem, these points are typically known and represent the position of an object in the world coordinate system. $\mathbf{R} \in \mathbb{R}^{3\times3}$ is a rotation matrix. $\mathbf{t} \in \mathbb{R}^{3\times1}$ is a translation vector. We can rewrite Equation (1) as Equation (2)

$$A_i \cdot (R|t) = u_i \tag{2}$$

where $\mathbf{A_i} \in \mathbb{R}^{4\times4}$ is a matrix constructed from the image point, f the focal length and the principal point coordinate (u_0, v_0). And $(\mathbf{R}|\mathbf{t})$ is an augmented matrix that includes the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$. For each corresponding point, we obtain an equation of this form. If there are n corresponding points, we can construct a linear system as Equation (3)

$$A \cdot (R|t) = U \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the matrix composed of $\mathbf{A_i}$, and $U \in \mathbb{R}^{n\times n}$ is the matrix composed of u_i. Finally, we can use SVD or other methods to solve this linear system, obtaining estimates for the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$. Then, we can use these estimates to calculate the focal length f.

IV. IMPLEMENTATION DETAILS

Our experimental platform is based on Ubuntu 20.04, CUDA 11.8, and an NVIDIA GeForce RTX 3090 GPU. We start with a learning rate of 0.01 and divide the learning rate by 10 at every 100 epochs. Model construction and training were carried out using the PyTorch framework.

During the training of our model, the choice of hyperparameters is crucial for its effectiveness. This section details the essential hyperparameters and standard settings. The initial learning rate, which sets the pace of model updates at the start of training, is commonly set to 0.01 or 0.001. The final learning rate, a key component in the learning rate schedule, is typically reduced to 20% of the initial rate. The momentum parameter, which ensures a smooth optimization process, is often set to 0.937. To prevent overfitting, weight decay is recommended at 0.0005. The warmup period, measured in epochs, allows the model to ease into training, typically lasting 3 epochs. The loss function weights for box, class, and objectness losses are balanced at 0.05, 0.5, and 1.0, respectively.

V. EXPERIMENTS

In this section, we will elaborate on the critical steps of the experimental procedure and introduce the datasets utilized. Through the description provided in this part, readers will gain a clear understanding of how the experiments were designed and conducted, as well criteria for dataset selection and their roles in the experiments.

A. DATASETS

COCO128 Datasets The COCO dataset serves as a standard benchmark for object detection and segmentation in complex, real-world scenarios. It comprises a vast collection of images with extensive annotations, including object bounding boxes, segmentation masks, and key points for a diverse array of object categories. The dataset's design facilitates the training of models to accurately identify and segment objects within various contexts, making it an invaluable resource for advancing the state of the art in computer vision. With over 200,000 images and over 1.5 million labeled instances, COCO provides a robust platform for evaluating and comparing the performance of object detection algorithms. Self-Built Datasets The self-established dataset is primarily divided into two segments: simulated data and real data. The simulated data are generated under ideal conditions, neglecting camera distortion errors and other factors, to serve as a benchmark for comparison with the estimation results. This data provides theoretical exact poses to facilitate the analysis and comparison with the actual measured outcomes. The real data are collected under the same arrangement and settings as the simulated data but incorporate various errors that may occur during the actual shooting process. By capturing images in diverse scenarios, these data more accurately simulate real-world pose estimation conditions, offering valuable information for the training and validation of algorithms.

B. DETAILED DESCRIPTION OF SELF-BUILT DATASETS

Our self-built dataset for this paper includes 3000 images, of which 2000 were collected under normal conditions and 1000 under challenging conditions like low light and obstructions. The dataset has been split, with 80% designated for the training set and 20% for the testing set, to ensure a thorough and balanced evaluation of our model.

C. EVALUATION METRICS

We employ five critical metrics: 1) mAP_{50} and 2) $mAP_{50:95}$ to evaluate the Enhanced Object Detection Network. We could get mAP_{50} from Equation (4), which represents the proportion of targets correctly detected by the model at a 50% detection accuracy threshold. And $mAP_{50:95}$ indicates the proportion of targets accurately detected within a range of 50% to 95% detection accuracy, which could be calculated from Equation (5).

$$mAP_{50} = \frac{1}{N}\sum_{i=1}^{N} AP_i(1, 0.5). \tag{4}$$

$$mAP_{50:95} = \frac{1}{N}\sum_{i=1}^{N} AP_i(1, [0.5, 0.95]). \tag{5}$$

where AP_i denotes the Average Precision for the $i-th$ category at an Intersection over Union (IoU) threshold of 0.5. N denotes the total number of categories. We use the mean square error (MSE) to evaluate pose estimation accuracy. The 3) MSE_x represents the mean square error (MSE) in X-axis orientation, which could be obtained from (6). 4) MSE_y represents the mean square error (MSE) in Y-axis orientation, which could be obtained from Equation (7).

$$MSE_x = \frac{1}{n}\sum_{i=1}^{n}(X_{i,pred} - X_{i,true})^2. \tag{6}$$

$$MSE_y = \frac{1}{n}\sum_{i=1}^{n}(Y_{i,pred} - Y_{i,true})^2. \tag{7}$$

These formulas, $X_{i,pred}$ and $X_{i,true}$ represent the predicted and actual X-axis values. $Y_{i,pred}$ and $Y_{i,true}$ represent the predicted and actual Y-axis values. $Z_{i,pred}$ and $Z_{i,true}$ represent the predicted and actual Z-axis values respectively. We denote the mean squared error(MSE) of the α angle by 5) MSE_α which could be obtained from Equation (8).

$$MSE_\alpha = \frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,pred} - \alpha_{i,true})^2. \tag{8}$$

while $\alpha_{i,pred}$ and $\alpha_{i,true}$ denote the predicted and actual angles, respectively.

D. ENHANCEMENT OF OBJECT DETECTION

In order to fairly compare the performance of different attention modules, we sought to integrate YOLOv5 with various attention modules to enhance feature representation capabilities. We evaluated the Squeeze-and-Excitation(SE) module and four attention modules: Efficient Channel Attention(ECA) [31], DoubleAttention(A2) [32], and Convolutional Block Attention Module(CBAM) [33] These modules possess distinct characteristics and advantages, and we incorporated them into the YOLOv5 backbone network to assess their performance in target detection tasks. We recorded the training loss, validation loss, and mAP metric for the model on the COCO128 dataset.

TABLE 1. Performance Metrics of YOLOv5 with Attention Modules.

Model	mAP_{50}	$\mathrm{mAP}_{50:95}$
YOLOv5	0.947	0.778
+ECA	0.971	0.845
+SE	**0.973**	**0.858**
+A2	0.971	0.818

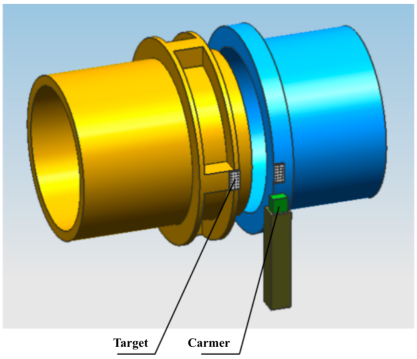

FIGURE 2. Cabins, target, and camera placement.

In assessing the YOLOv5 object detection model enhanced with an attention mechanism, we primarily focus on two evaluation metrics: mAP_{50} denotes the proportion of targets that the model can accurately detect when the detection accuracy is 50%; Similarly, $mAP_{50:95}$ signifies the proportion of targets that themodel can correctly detect when the detection accuracy ranges between 50% and 95%. Both of these evaluation metrics can provide a comprehensive assessment of the model's detection performance, particularly concerning detection accuracy. From Table 1, it is evident that the SE attention module demonstrates the most significant impact. The mAP_{50} value for YOLOv5+SE achieved 0.973. Concurrently, $mAP_{50:95}$the value stands at 0.858. These demonstrate the SE attention module in enhancing the performance of the YOLOv5 object detection model.

E. APPLICATION OF POSENOCAL-YOLOv5 IN TASK

Considering the issue of the cabin section's large volume, which surpasses the field of view of a standard camera, we have incorporated target key points as illustrated in Figure 3 on one side of the cabin section. To better differentiate between Cabin A and Cabin B, we have assigned red key points and blue key points respectively. Given that the cabin's interior space is a rigid entity, its geometric dimensions remain constant during movement, as do the relative positions between its internal points. Consequently, the attitude parameters of the external target object can be deduced from its pose parameters. As shown in Figure 3, includes two cabin sections (left and right), target key points, and the position of the monocular camera. The yellow cabin on the left is denoted as Cabin A, serving as a fixed section. The blue cabin on the right is denoted as Cabin B, representing the mobile end. The targets are attached to the

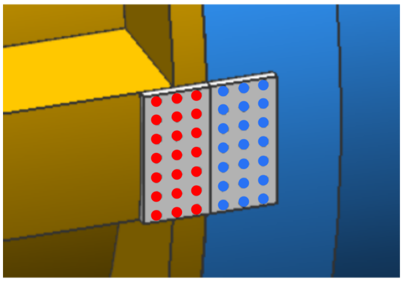

FIGURE 3. The key points placement.

TABLE 2. Retrained on Self- Built Database.

Model	mAP$_{50}$			mAP$_{50:95}$		
	All	RedA	BlueB	All	RedA	BlueB
YOLOv5	0.995	0.995	0.995	0.770	0.776	0.764
YOLOv5+ECA	0.995	0.995	0.995	0.825	0.832	0.818
YOLOv5+SE	**0.995**	**0.995**	**0.995**	**0.834**	**0.840**	**0.828**
YOLOv5+A2	0.995	0.995	0.995	0.812	0.825	0.799

edges of Cabin A and Cabin B. The monocular camera is positioned directly facing the targets.

Subsequently, we intend to apply this algorithm to our specific data set and retrain it. The result presented in Table 2. The mAP_{50} value remains stable at 0.995, and the $mAP_{50:95}$ value remains stable between 0.828 and 0.840, which indicates the effectiveness of the network. Our objective is to enhance the algorithm's performance by extracting key feature points from the data with greater accuracy through this process and to concurrently output detailed coordinates of the key points' centers.

Thereafter, we will utilize the output of these object center points as input data for pose estimation to ascertain the pose of the target object. To validate the effectiveness of our algorithm, we will meticulously compare these computed pose results with our self-built simulated database and calculate the corresponding covariance. In the context of evaluating an algorithm's suitability for practical application with respect to precision requirements, two critical criteria are typically employed 1)estimation precision: The mean square error in the X-direction should be controlled within 1mm, and the same applies to the Y-direction, while the mean square error for the alpha angle should be less than or equal to 0.05°. 2) Localization precision: The mean square error in the X-direction should be within 5 mm, and similarly for the Y-direction, with the mean square error being less than or equal to 0.1°.

As indicated by the results presented in Table 3, the MSE of measurement precision in the X-direction is 0.856mm less than 1mm. The MSE of measurement precision in Y-direction is 0.877mm less than 1mm. The MSE of measurement precision in angle is 0.048°less than 0.05°. The MSE of localization precision in the X-direction is 0.794mm less than 1mm. The MSE of measurement precision in Y-direction is 0.776mm less than 1mm. The MSE of measurement precision in angle is 0.869° less than 0.1°. The precision of our algorithm PoseNoCal-YOLOv5 meets the required standards, thereby validating its effectiveness in practical applications. This confirmation suggests that the algorithm is reliable and capable of performing as intended in real-world scenarios, which is crucial for its adoption and integration into various fields.

TABLE 3. MSE Results In Different Orientations.

	MES$_x$ (mm)	MES$_y$ (mm)	MES$_\alpha$ $(^\circ)$
Estimation	0.856	0.877	0.048
Localization	0.794	0.776	0.869

VI. CONCLUSION

In this paper, we introduced the PoseNoCal-YOLOv5 algorithm, an innovative method specifically designed for missile docking pose estimation. The algorithm not only demonstrates excellence in pose estimation but also achieves significant performance improvements in the task of target detection. Specifically, the incorporation of attention mechanisms into YOLOv5 led to a substantial 2.6% increase in $mAP_{50:95}$, a key metric for target detection in coco128 datasets. And 6.4% increase in $mAP_{50:95}$ on the self-built datasets. For pose estimation, our method demonstrated an average error reduction of 14.4% on the X-axis and 12.3% on the Y-axis, both surpassing the project(Science and Technology Development Program Project of Jilin Province)'s 1mm error tolerance. Additionally, our algorithm achieved a 2% reduction in angular measurement error, meeting the project's precision criteria.

These enhancements not only demonstrate the capability of the PoseNoCal-YOLOv5 algorithm in high-precision pose estimation but also showcase its reliability and effectiveness in practical applications. Compared to existing methods, PoseNoCal-YOLOv5 has shown significant performance improvements across multiple key performance indicators. While the initial results are auspicious, we recognize that there is still room for further optimization and refinement. Future work will focus on these areas to enhance the performance and applicability of the PoseNoCal-YOLOv5 algorithm. We are committed to continuous optimization to ensure that our algorithm meets the evolving demands of practical applications and remains at the forefront of technological advancements in the field of pose estimation. The successful application of the PoseNoCal-YOLOv5 algorithm in missile pose estimation has set the stage for future development to concentrate on integrating state-of-the-art technologies from various domains. Particularly, the incorporation of high-precision sensor chips is expected to enhance the precision and responsiveness of missile guidance systems. The miniaturization and integration of these sensor chips are anticipated to bolster the algorithm's real-time data processing capabilities. Furthermore, this could catalyze the automation and intelligence of missile systems, providing innovative solutions for the aerospace and defense industries.

REFERENCES

[1] R. Siegers, ''The Raytheon enterprise architecture process (REAP),'' in *Proc. INCOSE Int. Symp.*, 2003, vol. 13, no. 1, pp. 1229–1240.

[2] G. Peng, M. Ji, Y. Xue, and Y. Sun, ''Development of a novel integrated automated assembly system for large volume components in outdoor environment,'' *Measurement*, vol. 168, Jan. 2021, Art. no. 108294.

[3] H. B. Wang, Y. Li, K. Ren, L. J. Yang, and Z. H. Han, ''The development status and trends of ground unmanned combat platforms,'' *J. Phys., Conf. Ser.*, vol. 1721, no. 1, Jan. 2021, Art. no. 012065.

[4] C. Cheng, X. Li, L. Xie, and L. Li, ''Autonomous dynamic docking of UAV based on UWB-vision in GPS-denied environment,'' *J. Franklin Inst.*, vol. 359, no. 7, pp. 2788–2809, May 2022.

[5] Boeing Company, ''An autonomous docking system for use in Boeing 787 dreamliner and other aircraft,'' U.S. Patent 13 072 295, Oct. 26, 2011.

[6] Y. Zhang, Y. Li, and G. Zhang, ''Development of automatic docking system for A380 aircraft,'' *Aerosp. Sci. Technol.*, vol. 95, Jan. 2019, Art. no. 107955.

[7] G. Jocher. (May 21, 2020). *YOLOv5: PyTorch Implementation of YOLO*. [Online]. Available: https://github.com/ultralytics/yolov5

[8] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, ''Squeeze-and-excitation networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[9] A. Penate-Sanchez, J. Andrade-Cetto, and F. Moreno-Noguer, ''Exhaustive linearization for robust camera pose and focal length estimation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2387–2400, Oct. 2013.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[12] K. Simonyan and A. Zisserma, ''Very deep convolutional networks for large-scale image recognition,'' 2015, *arXiv:1409.1556*.

[13] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[15] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[16] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards real-time object detection with region proposal networks,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[18] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, ''Mask R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2980–2988.

[20] Z. Cai and N. Vasconcelos, ''Cascade R-CNN: Delving into high quality object detection,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.

[21] Z. Tian, C. Shen, H. Chen, and T. He, ''FCOS: Fully convolutional one-stage object detection,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635.

[22] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, ''Complete solution classification for the perspective-three-point problem,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003.

[23] L. Ferraz, X. Binefa, and F. Moreno-Noguer, ''Leveraging feature uncertainty in the PnP problem,'' in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 83.1–83.13.

[24] S. Urban, J. Leitloff, and S. Hinz, ''MLPnP—A real-time maximum likelihood solution to the perspective-*n*-point problem,'' 2016, *arXiv:1607.08112*.

[25] Y. Wu, F. Tang, and H. Li, ''Image-based camera localization: An overview,'' *Vis. Comput. Ind., Biomed., Art*, vol. 1, p. 8, Sep. 2018.

[26] M. W. Cao, W. Jia, and Y. Zhao, ''Fast and robust absolute camera pose estimation with known focal length,'' *Neural Comput. Appl.*, vol. 29, pp. 1383–1398, Mar. 2018.

[27] Q. Yu, G. Xu, L. Zhang, and J. Shi, ''A consistently fast and accurate algorithm for estimating camera pose from point correspondences,'' *Measurement*, vol. 172, Feb. 2021, Art. no. 108914.

[28] A. Vakhitov, L. F. Colomina, A. Agudo, and F. Moreno-Noguer, ''Uncertainty-aware camera pose estimation from points and lines,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4657–4666.

[29] G. Nakano, ''Efficient DLT-based method for solving PnP, PnPf, and PnPfr problems,'' *IEICE Trans. Inf. Syst.*, vol. E104.D, no. 9, pp. 1467–1477, 2021.

[30] Y. Zhang, Y. Zhang, B. Hu, Y. Yin, W. Chen, X. Liu, and Q. Yu, ''An efficient and accurate solution to camera pose estimation problem from point and line correspondences based on null space analysis,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 3762–3769.

[31] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, ''ECA-Net: Efficient channel attention for deep convolutional neural networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[32] Y. Chen, Y. Kalantidis, J. Li, and J. Feng, ''A^2-Nets: Double attention networks,'' in *Proc. Neural Inf. Process. Syst.*, 2018, pp. 1–10.

[33] S. Woo, J. Park, J. Lee, and I. S. Kweon, ''Synthetic structure of industrial plastics,'' in *CBAM: Convolutional Block Attention Module*, vol. 11211. Cham, Switzerland: Springer, 2018, pp. 3–19.

[34] G. Shi, X. Shen, F. Xiao, and Y. He, ''DANTD: A deep abnormal network traffic detection model for security of industrial Internet of Things using high-order features,'' *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21143–21153, Dec. 2023, doi: 10.1109/JIOT.2023.3253777.

[35] G. Shi, X. Shen, Y. He, and H. Ren, ''Passive wireless detection for ammonia based on 2.4 GHz square carbon nanotube-loaded chipless RFID-inspired tag,'' *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 9510812, doi: 10.1109/TIM.2023.3300433.

[36] X. Shen, G. Shi, H. Ren, and W. Zhang, ''Biomimetic vision for zoom object detection based on improved vertical grid number YOLO algorithm,'' *Frontiers Bioeng. Biotechnol.*, vol. 10, May 2022, Art. no. 905583.

[37] X. Shen, G. Shi, Y. Zhang, and S. Weng, ''Wireless volatile organic compound detection for restricted Internet of Things environments based on cataluminescence sensors,'' *Chemosensors*, vol. 10, no. 5, p. 179, May 2022.

[38] S. Liu, Y. Jin, and H. Tang, ''Design of the automatic docking system of missile cabin,'' *Modular Mach. Tool Autom. Manuf. Technique*, no. 2, pp. 103–106, 2022.

[39] X. Cao, D. Mao, C. Liu, S. Zhou, and K. Zhu, ''Design of integrated platform for flexible assembly and measurement of cabin,'' *Ind. Instrum. Autom. Devices*, no. 2, pp. 30–38, 2024.

CHANGHONG JIANG received the Ph.D. degree in agricultural machinery design and manufacturing from Jilin University. He is currently the Dean of the Graduate School and a Ph.D. Supervisor with Changchun University of Technology. He has authored more than 55 academic articles and secured more than 50 invention patents, contributing significantly to his fields of research. His research interests include multiscale modeling and control of complex systems, and machine signal detection and information processing. He has been awarded the first prize in scientific and technological progress by China Petroleum and Chemical Automation Industry, and one second prize and one third prize in scientific and technological progress by Jilin Province. He is a recipient of the 8th Jilin Province Youth Science and Technology Award, the Principal Investigator of a project under the National Science and Technology Support Plan, a Distinguished Innovator in the 4th Batch of Top-Tier Talents in Jilin Province (third level), an Expert with outstanding contributions in the 5th Batch of Changchun City, and one of the ''Top 100 Outstanding Scientists and Technicians'' in the 2nd Changchun City Selection.

XIAOQIAO MU received the master's degree in electrical and computer engineering from Portland State University, in 2020. She is currently pursuing the Doctor of Mechanical Engineering.

BINGBING ZHANG received the Ph.D. degree in signal information and processing from Dalian University of Technology, in 2023. She is currently a Lecturer with the School of Computer Science and Engineering, Dalian Minzu University. Her work in computer vision has resulted in 15 publications. Her research interests include computer vision, deep learning, and video action recognition.

MUJUN XIE received the Ph.D. degree in optical engineering from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China, in 1999. Her main research interests include industrial machinery fault diagnosis technology and intelligent machinery and robot control. Aiming at the machinery, such as synchronous motor, and air compressor which are widely used in modern industry, based on multi-sensor information fusion technology, the research on obstacle signal acquisition, processing, fault diagnosis, and location is carried out to achieve fault prediction and judgment. Aiming at intelligent machinery and robot, multi-sensor information fusion vision measurement, system modeling, and control strategy research are carried out to realize the robot position and attitude control.

CHAO LIANG received the master's degree in signal and information processing from Changchun University of Technology, in 2007. He has been working there ever since. He has published more than ten articles in various journals. His main research interests include machine vision, deep learning, and intelligent control.

• • •