Improved YOLOv11pose for Posture Estimation of Xinjiang Bactrian Camels

Lei Liu, Alifu Kurban,Yi Liu
Department School of Software, Xinjiang University, Urumqi, China

***Abstract*—Automatic pose estimation of camels is crucial for long-term health monitoring in animal husbandry. There is currently less research on camels, and our study has certain practical application value in actual camel farms. Due to the high similarity of camels, this has brought us a huge challenge in pose estimation. This study proposes YOLOv11pose-Camel, a pose estimation algorithm tailored for Bactrian camels. The algorithm enhances feature extraction with a lightweight channel attention mechanism (ECA) and improves detection accuracy through an efficient multi-scale pooling structure (SimSPPF). Additionally, C3k2 modules in the neck are replaced with dynamic convolution blocks (DECA-blocks) to strengthen global feature extraction. We collected a diverse dataset of Bactrian camel images with farm staff assistance and applied data augmentation. The optimized YOLOv11pose model achieved 94.5% accuracy and 94.1% mAP@0.5 on the Xinjiang Bactrian camel dataset, outperforming the baseline by 2.1% and 2.2%, respectively. The model also maintains a good balance between detection speed and efficiency, demonstrating its potential for practical applications in animal husbandry.**

Keywords—YOLOv11pose; efficient channel attention; multi-scale pooling structure; DECA-block; Bactrian camel posture estimation；SimSPPF; ECA

I. Introduction

With the continuous large-scale development of the modern camel farming industry, their farming methods have gradually changed from traditional free-range farming to cluster farming. The cluster farming method makes the breeders pay more and more attention to the healthy development of camels. Breeders increasingly recognize the critical role of health management in selecting and cultivating high-quality camel breeds. As an essential metric of health management, posture estimation is gaining growing attention. It provides scientific data for complex behavior analysis and health monitoring tasks, extending beyond the scope of traditional detection methods and better adapting to diverse scenarios. By estimating the key points of camels, we can further analyze their behaviors. The external behaviors of camels often reflect their health conditions. For instance, daily postures such as standing or walking can indicate their activity level, as camels tend to reduce activity when ill [1]. However, long-term manual observation and recording of individual information are time-consuming, costly, and subjective. Currently, many researchers use various sensors to monitor animal behavior [2-5]. Nevertheless, wearable sensors have limitations, as they may stress the animals and impact their natural growth and development. In contrast, machine vision offers long-term, non-contact continuous monitoring and has gradually been applied in precision livestock farming to monitor livestock activities and health conditions. With the rapid advancements in image processing [6], pattern recognition [7], and artificial intelligence [8], machine learning applications in China's animal husbandry are also increasing. Machine vision-based camel posture detection facilitates comprehensive analysis and rapid assessment of camel growth and development [9].

The rapid development and widespread application of computer vision technology provide robust technical support for the intelligent, precise, and scientific modernization of livestock farms. The mature implementation of these technologies has laid a solid foundation for the advancement of modern animal husbandry [10]. Through object detection techniques and keypoint detection algorithms, it is possible to extract the key features of animal postures, which can then be connected to form skeletal structures, enabling accurate posture estimation [11]. In the task of pose estimation, our focus is on ensuring the accuracy of keypoint detection. During the detection process, we often encounter various complex problems that can affect the accuracy of keypoint detection. For example, camels have relatively large bodies, and diverse poses, and tend to move in close groups. These problems can cause key points on certain parts of camels to be obscured by other camels or objects. In addition, the deformation of features on camels in different poses can cause key points to blur, thereby affecting the model's detection of skeletal key points. These problems not only affect the accuracy of keypoint localization but also the precision of detection boxes. If the skeletal features extracted from the image are incomplete, accurate pose estimation cannot be performed.

To address the aforementioned challenges, we have proposed a Bihumped Camel pose estimation framework, named YOLOv11pose-Camel, which is designed to effectively process camel information in various forms and environments. Specifically, we first utilize channel attention and multi-scale pooling to enhance sensitivity to detailed features. Secondly, the error under complex postures is further reduced by incorporating our DECA-Block module. The main contributions of this paper are as follows:

1) We used data augmentation techniques to cover a variety of environments with camels in the dataset and applied the improved YOLOv11-pose algorithm to our self-built dataset.

2) We introduced the ECA lightweight channel attention mechanism into the model to enhance the algorithm's multi-scale feature extraction ability. The model effectively reduced the extraction of non-camel features and focused more on the precise identification of camel skeleton points [12].

3) To further enhance the algorithm's ability to identify skeleton key points, we combined the algorithm with the efficient multi-scale pooling structure SimSPPF. This can effectively reduce the precision error caused by occlusion.

4) We introduced the improved dynamic convolution module DECA-Block into the neck module of the YOLOv11-pose model. This module combines dynamic convolution and an efficient attention mechanism. This design not only reduces the model's weight but also maintains a high accuracy. In particular, the module shows strong adaptability to feature deformation problems.

The algorithm adopted in this study can meet the actual needs for detecting camel skeleton key points. Meanwhile, the method can infer the information of occluded skeleton points. Section II of the paper reviews the current research status on Bactrian camel posture estimation both domestically and internationally, as well as its application in animal husbandry. Section III describes the overall network structure and focuses on the proposed YOLOv11pose-Camel method, which integrates the ECA channel attention mechanism, the multi-scale feature module SimSPPF, and the dynamic convolution module DECA-Block. Section IV introduces the experimental dataset, evaluation metrics, and recognition performance, followed by comparative analyses and experimental results on the custom dataset. Finally, Section V summarizes the main contributions and outlines potential future research directions.

II. Related Work

In recent years, pose estimation based on keypoint detection has significantly advanced image recognition and analysis technologies, achieving notable progress in animal posture detection. In early studies, Fang et al. (2017) introduced a novel multi-person pose estimation framework called RMPE, aimed at enhancing the adaptability of deep convolutional networks to inaccuracies in bounding box localization. This innovation not only improved feature extraction efficiency but also laid the foundation for subsequent developments in multi-person pose estimation. However, since RMPE was primarily designed for humans and camels possess unique body shapes and complex keypoint distributions, it cannot effectively estimate camel poses [13]. In 2018, Zheng et al. incorporated Faster R-CNN into a deep learning framework for pig posture estimation, enabling the detection of five posture changes (standing, sitting, sternal recumbency, ventral recumbency, and lateral recumbency) to reflect the health status of sows [14]. However, the more drastic posture variations in camels make it challenging for a single Faster R-CNN framework to capture their dynamics. Similarly, Song et al. (2018) proposed a model for skeletal extraction based on keypoint prediction in walking states, which performed well in detecting cow walking postures [15]. Comparatively, research focused directly on camel posture estimation remains limited. In 2020, Chen et al. utilized the real-time instance segmentation framework YOLACT for pig body part tracking [16]. However, camels' unique physiological structures, especially the presence of humps, add significant challenges in segmentation and localization. In 2023, Zheng et al. developed the ViT-BERT pose estimation model for rapid posture changes in small animals within dynamic scenes. While this model demonstrated good accuracy and speed in bird testing, it is less suitable for large animals like camels [17]. Also in 2023, Natesan et al. proposed a YOLOv5-based algorithm incorporating adaptive attention mechanisms to estimate occluded key points, improving measurement accuracy and achieving precise livestock behavior recognition [18]. In 2023, Agullo and colleagues combined MobileNetV3 with Vision Transformer (ViT) for animal pose detection, aiming to perform real-time pose estimation with lower computational resources. This approach aligns well with the current needs of automated livestock farms [19]. However, for large animals with complex postures, higher resolution and deeper feature extraction are required, and resource constraints may limit accuracy. In 2022, Wang et al. introduced a dynamic convolution and bidirectional LSTM time series ensemble, which considered long-term posture tracking and showed good performance in complex groups [20]. However, this method has not been fully explored in applications where posture changes frequently or under occlusion conditions. In 2023, Zhao and his colleagues enhanced their model by combining ResNet with LSTM for behavior recognition and pose detection, allowing for the identification of abnormal situations while recognizing poses. This algorithm primarily applies to animals with uniform body types, and the complexity and specificity of camel bodies pose significant challenges for applying this technology to camel pose detection [21]. In 2023, Barney and his research team used deep learning methods for cattle detection and pose estimation. They modified Mask-RCNN to estimate poses in video sequences for each cow and applied the CatBoost gradient boosting algorithm to combine all features, using triple cross-validation to determine accuracy. This method proved highly effective for detecting hoof issues in cattle, which is critical for animal health. However, the same measurement point localization is not suitable for camels, as their body structure is more complex [22]. Additionally, in 2024, Dhivya Mohanavel and others proposed an animal detection early warning system based on deep learning and computer vision technologies for the livestock environment, further highlighting the growing importance of these technologies in the agricultural industry [23]. This research offers valuable insights into the practical application of animal management in real-world agricultural settings; however, it has not been thoroughly explored in the context of precise analysis of bipedal pose characteristics.

The aforementioned research demonstrates the effectiveness of non-contact monitoring methods and highlights significant progress in this field. In the future, non-contact computer vision methods are likely to become mainstream in research, not only effectively reducing labor costs but also promoting animal welfare. Existing technologies have shown certain success in animals such as humans, cattle, and pigs, but camel pose estimation still faces numerous challenges. Due to camels' unique body structure, gregarious nature, and variability, pose estimation becomes particularly difficult. With the development of computer vision technologies [24], there is potential for further research on optimization models focused on camels, which could better support their morphological evaluation and health monitoring. This would not only enhance the practical value of the model but also provide valuable insights for the application of deep learning and computer vision in agriculture.

III. Research Methodology

The YOLOv11-pose model is a deep convolutional network based on YOLOv11, designed with an end-to-end unified structure rather than the traditional two-branch composition commonly used in pose estimation. This single-stage network performs both keypoint localization and partial affinity field prediction simultaneously, enabling pose estimation through joint training of object detection and keypoint detection. Initially developed for human pose estimation, the model has demonstrated promising results in animal pose estimation as well. The YOLOv11pose-Camel model consists of four main components: input, backbone, neck, and head. The lower-level features provide detailed information about object positions, while higher-level features capture stronger semantic information. In the input stage, multiple convolutional layers (Conv) preprocess the input image to the dimensions required for model training and extract preliminary features. The backbone includes several C3K2 modules, which merge features from different layers and pass the refined features to the SimSPPF module for further multi-scale processing. Subsequently, the ECAttention module enhances the feature extraction process by emphasizing attention mechanisms, capturing the relational information between different spatial regions through feature fusion. This allows the model to focus on critical feature regions, improving detection performance and convergence speed. In the neck, the PAN structure is employed [25], and the DECA-Block module is integrated to enhance the network’s capability in multi-scale feature integration. The head network adopts depthwise separable convolution to process the network's outputs and convert them into the required format for the output layer. This design reduces computational overhead without compromising keypoint accuracy, ensuring model stability. The proposed model architecture is illustrated in Fig. 1.

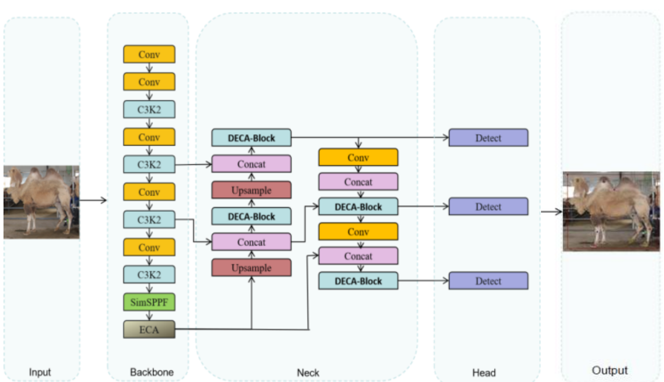

Fig. 1. Architecture of the YOLOv11pose-camel model.

A. ECAttention Mechanism

In deep convolutional neural networks, the primary function of channel attention mechanisms is to enhance the model's focus on significant features, enabling more precise feature extraction and improved performance. As input images pass through convolutional networks, spatial information is gradually embedded into channels. However, repeated spatial compression or channel expansion may lead to the loss of some semantic information. By dynamically assigning higher weights to important features, the model can concentrate more effectively on task-relevant features, thereby enhancing accuracy [26]. Thus, channel attention mechanisms are critical in deep convolutional neural networks. To achieve a balance between detection speed and efficiency, we chose to employ the lightweight Efficient Channel Attention (ECA) mechanism [27]. ECA improves upon the SENet architecture by removing fully connected layers and replacing them with a 1*1 convolution kernel. This introduces a local cross-channel interaction strategy with reduced dimensionality, effectively meeting the requirements for balancing detection accuracy and speed. Specifically, the ECA module adaptively determines the kernel size k for one-dimensional convolution based on the number of input channels. The calculation formula is as follows:

$$F = CNN(I)[k = \left| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}] \tag{1}$$

Where C represents the number of channels for the input feature, γ and b is the hyperparameter. The core idea of the ECA module is to introduce a one-dimensional convolution operation into the relationship between input feature channels after determining kernel size k. The ECA attention mechanism first performs global average pooling operations on the input feature maps to generate global information. Then, the module performs one-dimensional convolution processing with kernel size k on the globally pooled information. The activation function maps the weights nonlinearly to obtain the weights for each channel. Finally, these weights are multiplied with the original input feature graph channels to produce output features that incorporate the attention weights. The detailed flow of the attention mechanism is shown in Fig. 2.

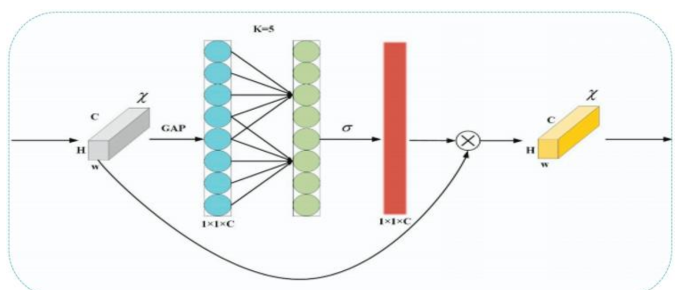

Fig. 2. ECAttention workflow diagram.

B. SimSPPF

SimSPPF (Simplified Space Pyramid Pooling Fast) is an optimized version of the fast version of the space pyramid pooling (SPPF) based on the classic space pyramid pooling (SPP). SPP is an improved version of the classic space pyramid pooling. The feature map is pooled at multiple scales to improve the receptive field in SPP. Although this method has significant effects, the computational cost is high. SPPF reduces redundant content, not only improving the efficiency of feature extraction but also reducing memory usage. The main innovation of SimSPPF is the hierarchical structure, which divides nodes into different layers according to their size, and the number of nodes in each layer is twice that of the previous layer. SimSPPF can reuse allocated nodes, which reduces the computational cost. In addition, this improvement reduces the dimensionality of the feature network. Specifically, SimSPPF replaces the activation function in the original SPPF with a new activation function (ReLU). This combination of multi-scale pooling methods not only captures multi-level feature information but also ensures a certain detection ability in different scenarios. Fig. 3 shows the structure design of SPP, and Fig. 4 shows SPPF and SimSPPF.

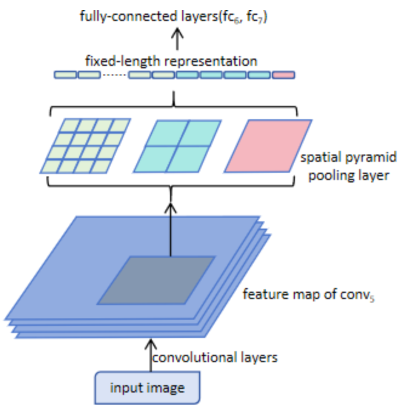

Fig. 3. Structure of SPP.

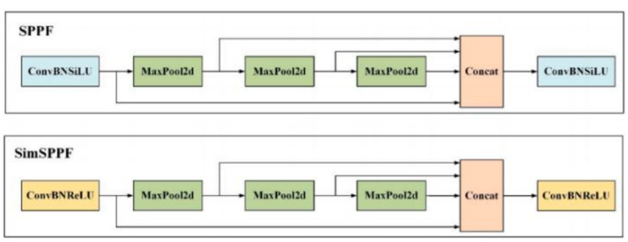

Fig. 4. Structures of SPPF and SimSPPF.

C. DECA-Block

The neck module primarily optimizes and fuses the features from the backbone network, improving the performance of key point detection tasks. When handling features of varying scales and backgrounds, the neck module addresses issues such as feature variation and keypoint occlusion, particularly in dealing with multi-scale features [29].

To enhance the model's understanding of feature details and structures, inspired by feature extraction components, we propose a dynamic convolution module called DECA-Block, designed for keypoint detection of various parts of Xinjiang Bactrian camels. This method helps mitigate keypoint localization errors and feature deformation among camels, adapting effectively to complex environments to balance model accuracy and efficiency. Specifically, the module incorporates the efficient feature extraction capability of the EMAttention mechanism and the deformable convolution network (DCNv3) into the bottleneck structure. EMAttention focuses on critical features and guides their transmission to DCNv3, which adapts to feature variations spatially. This enables the model to better capture key points across different scales and semantic levels, thereby improving detection accuracy. Such direct interaction enhances the attention mechanism's performance on deformed features, providing distinct advantages for complex shapes and occluded key points. Additionally, it reduces redundant computation and improves overall model performance and efficiency. The specific structure of the DECA-Block module is shown in Fig. 5.

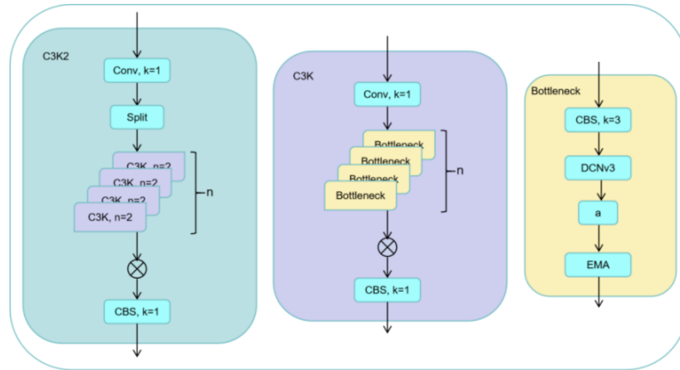

Fig. 5. D3E-C3K2 module.

D. Bactrian Camel Pose Estimation Method

The algorithm analyzes the three-dimensional joint position information contained in the two-dimensional image to achieve the analysis of the target pose. Based on this, the algorithm fully utilizes contextual information to solve the problem of joint overlap. Contextual information includes precise detection of joint positions, modeling of camel pose through structured constraints, and estimation of joint relative position relationships through analysis. Each object possesses its unique anatomical structure, and these structural constraints provide critical prior knowledge to the model. In cases where certain joints are occluded, this prior knowledge effectively aids the model in achieving more accurate pose estimation [30].

$$\mathrm{J} = \{\mathrm{j}_1, \mathrm{j}_2, \dots, \mathrm{j}_\mathrm{n}\} \tag{2}$$

Where j_n is the position vector of the n-th joint. The structure of the camel pose can be represented by a set of distance constraints between joints.

$$D_{ij} = \| j_i - j_i \|, \forall i, j \in \{1,2,\dots,n\} \tag{3}$$

These distance constraints D_i can be learned from the training data. During the pose estimation process, they are widely used to help the model infer the positions of joints that are not directly observable due to occlusion. YOLOv11-pose extracts feature maps from input images through convolutional neural networks (CNN). Assuming the input image is represented as i, the feature maps extracted by the CNN are denoted as F . These feature maps provide multi-scale structured information, laying the foundation for subsequent pose estimation tasks.

$$\mathrm{F} = \mathrm{CNN}(\mathrm{I}) \tag{4}$$

The feature maps F contain multi-scale features and rich contextual information of the image, which play a vital role in predicting the positions of skeletal joints. During the prediction of joint locations, the model relies not only on local features but also heavily on global features derived from the overall structure. The rich contextual information further provides comprehensive support. Additionally, the model leverages multi-level information to accurately analyze pose characteristics. For instance, for certain occluded skeletal key points j_k, the model can utilize the information from other detected key points in the feature map and infer the position by combining global contextual features. Annotated data sets are fed into the model for training, preparing it for subsequent pose estimation tasks. The annotations for Bactrian camels are illustrated in Fig. 6.

$$j_k = f(F, j_1, j_2, \dots, j_{k-1}, j_{k+1}, \dots, j_n) \tag{5}$$

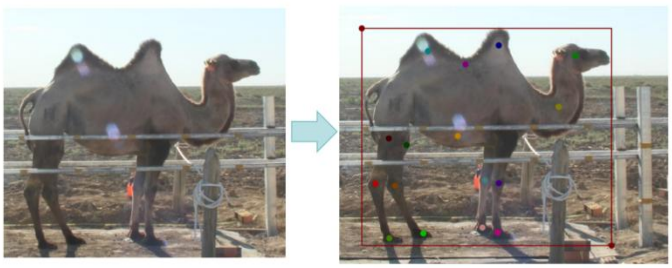

Fig. 6. Annotation of Bactrian camels.

IV. Preparation and Analysis

A. Dataset Preparation

In this study, a Canon 60D camera equipped with an 18-135mm lens was used for data collection. The initial image capture took place in December 2020 in Ziniquanzi Town, Fukang City, Xinjiang Uygur Autonomous Region, where images of Xinjiang Bactrian camels were recorded. Later, in March and April 2021 and July 2021, further collection was carried out, which included various lighting conditions, different angles and distances, and diverse obstruction scenarios, resulting in the acquisition of 90 high-quality images. Considering the limited dataset size, additional data collection was conducted in September 2023 in Keeping County, located in the western Aksu Prefecture of Xinjiang, a region renowned as the “Camel City of Xinjiang.” This supplementary collection carried out in collaboration with the local Xintuomilk Group, added approximately 252 effective images to the dataset.

1) Images selection: We constructed a dataset containing 1,084 images for the study of Xinjiang Bactrian camel pose estimation. The data collection was conducted across multiple scenarios and seasons, resulting in images captured under varying lighting conditions influenced by changes in sunlight.

2) Key points selection: We utilized the open-source image annotation tool Labelme [31] for manual annotation of the images. During the annotation process, we focused on creating bounding boxes around Bactrian camels to minimize the area of detection boxes and reduce the influence of unrelated factors. Due to the unique structure of Bactrian camels, we annotated 1 rectangular box ("camel") and 16 key points (A, B, C...P), representing the head, neck, first hump, back, second hump, tail, right hind knee, right hind ankle, right hind hoof, abdomen, right foreleg knee, right foreleg hoof, left foreleg knee, left foreleg hoof, left hind knee, left hind ankle, and left hind hoof. Invisible skeletal key points were labeled as "1." An example of the key point annotations is shown in Fig. 1. These annotations were saved as text files, corresponding to the image filenames, providing a solid foundation for subsequent training.

3) Data augmentation: We introduced various data augmentation techniques [32] in the training set to enhance the model's robustness in different scenarios and its adaptability to complex environments. These techniques included adding random noise, simulating varying levels of occlusion, and performing operations such as mirroring, flipping, rotating, cropping, and stitching at random locations. These methods effectively simulated real-world scenarios the model might encounter. The model not only provides a reliable scientific basis and data support for Bactrian camel morphological evaluation based on pose estimation but also ensures accuracy and comprehensiveness in the evaluation process. Data augmentation examples are shown in Fig. 7.

4) Dataset formatting: To ensure the dataset's representativeness and the reliability of model training, we randomly sampled images collected at different times. These samples were divided into three parts: the training set, validation set, and test set. The training and validation sets comprised 80% of the total dataset, with a 9:1 split between them. The remaining 20% was allocated to the test set for final model evaluation. Special attention was given to ensuring that the collection times of images in the test set differed from those in the training set, effectively reducing potential interference caused by temporal sequence effects on model performance evaluation.

Fig. 7. Results of data augmentation.

B. Main Hardware Equipment

The data collection equipment includes a Canon 60D camera, an 18-135mm lens, and a stereo camera with frame synchronization. For model training, the computational hardware utilized an NVIDIA GeForce RTX 4090 GPU, with PyTorch version 2.1.0 and Python version 3.10. The GPU computation platform was based on CUDA version 12.1. The model's input size was set to 640×640.

C. Evaluation Metrics

The evaluation metrics for Camel-YOLOv11pose include Mean Average Precision (mAP50), Mean Average Precision at thresholds ranging from 0.50 to 0.95 (mAP50-95), Precision

(P), Recall (R), and Parameters (Params). The mAP50-95 metric evaluates the average precision at different IoU thresholds from 0.5 to 0.95, providing a more comprehensive assessment of the model's performance [33].

D. *Experimental Results and Analysis*

1) Comparative experiments: We conducted comparative experiments to verify the performance of the proposed model in the pose estimation task and compared it with several classic pose estimation algorithms across multiple metrics. To ensure the scientific rigor of the experiment, we adopted the same strategy as in the ablation experiments. All experiments used the same evaluation metrics and were tested on the same dataset.

TABLE I. MODEL COMPARISON RESULTS

Camel	P	R	mAP50	mAP50-95	Parameters	FPS
YOLOv7n-pose	90.2	89.6	87.7	68.5	34930376	140.8
YOLOv8n-pose	91	94.3	90.5	77.9	2971999	163.9
YOLOv8m-pose	90	94.2	89.8	75.9	23820511	181.8
YOLOv8s-pose	91.4	95.1	91.7	78.3	10324383	144.9
YOLOv11n-pose	92.4	90.7	91.9	79.6	2869783	166.6
YOLOv11pose-Camel	94.5	94.3	94.1	82.8	2678642	147.6

As shown in Table I, our YOLOv11pose-Came model performs better in the camel pose estimation task for the Bactrian camels of Xinjiang. Although both YOLOv7-pose and YOLOv8-pose used pre-trained weights as initial settings, their keypoint prediction [34] accuracy did not reach the baseline model's level. While the FPS slightly decreased compared to the original YOLOv11-pose baseline model, other metrics were improved. Compared to YOLOv7n-pose, YOLOv8n-pose, YOLOv8m-pose, YOLOv8s-pose, and YOLOv11n-pose, the improved model's mAP50 increased by 6.4, 3.6, 4.3, 2.4, and 2.2 percentage points, respectively. Although YOLOv8m-pose has a higher FPS than the improved model, its larger parameter size leads to higher memory consumption. Therefore, this study effectively improves the accuracy of camel pose estimation for Xinjiang Bactrian camels while ensuring detection speed, validating the effectiveness of the proposed method. The model comparison results are shown in Table I [28].

2) Ablation experiments: Based on the current YOLOv11pose-Camel model, we designed ablation experiments to evaluate the enhancement effects of the proposed modules and their contributions to algorithm performance. These experiments aim to analyze the impact of individual modules in depth. First, we introduced the ECAttention mechanism into the backbone feature extraction network to enhance the network's focus on key points. Next, we replaced the original SPPF module in the existing feature extraction network with the SimSPPF module. Finally, we integrated the C3K2 module with the DECA-Block module, which combines the EMAttention mechanism and DCNv3 dynamic convolution. We conducted experimental evaluations of each improvement strategy on the same dataset. To ensure the accuracy of the results, we used the same evaluation metrics to measure the contribution of each improvement to the network. From these experiments, it can be observed that each module contributes to varying degrees of improvement in the overall performance of the model. Through the pose estimation algorithm, we successfully obtained accurate skeletal key points. The relevant experimental details and test results are shown in Table II.

TABLE II. ABLATION EXPERIMENT RESULTS

	P	R	mAP50	mAP50-95
YOLOv11-pose	92.4	90.7	91.9	79.6
ECAttention	93.3	91.9	92.5	81.7
SimSPPF	92.9	90.8	93.4	79.7
SimSPPF-ECA	93.8	92.2	93.3	82.5
YOLOv11-pose-Camel	94.5	94.3	94.1	82.8

3) Bactrian camel pose estimation: We determined the position information of the key points by using the pose estimation algorithm, and then inferred the occluded key points using Eq. (5). These key points allow for a more accurate description of the camel's skeletal structure and dynamic changes, providing scientific data for subsequent studies on the morphology of Bactrian camels [35]. Some training results of YOLOv11pose-Camel are shown in Fig. 8.

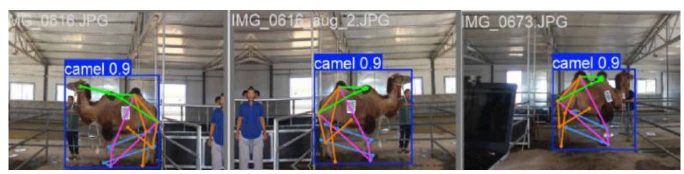

Fig. 8. Partial training results.

V. SUMMARY AND OUTLOOK

A. *Summary*

In this study, we explored the pose estimation method for Bactrian camels and conducted detailed technical analysis and experimental validation. By introducing the YOLOv11pose-Camel optimization design, we successfully achieved an efficient fusion of object detection and pose estimation, improving recognition accuracy and real-time performance under complex backgrounds. The model significantly enhanced the ability to recognize skeletal key points of different parts of Bactrian camels. YOLOv11pose-Camel maintained good.

Performance under various experimental conditions, achieving evaluation metrics (precision, recall, mean average precision) of 94.5%, 94.3%, and 94.1%, respectively. These performances surpassed existing models such as YOLOv8pose, YOLOv7pose, YOLOv5pose, and the base model YOLOv11pose used in this study. The improved model reduced parameters by 0.191 million, and the FPS reached 158.73, reflecting a significant improvement in detection speed and real-time performance, making it highly suitable for rapid deployment and cross-device portability.

In the context of the current development of intelligent farming, pose estimation and real-time monitoring of Bactrian camels are becoming increasingly important. Pose estimation lays the technical foundation for subsequent morphological evaluation (such as hoof health and rumination), which can ensure the healthy development of camels at an early stage, providing valuable insights into enhancing economic value. Therefore, the results of this study hold significant appeal for camel breeders [36], [29].

B. Outlook for the Future

Although this study has made certain progress in some areas, there are still some limitations.

1) This research focuses on pose estimation for Bactrian camels in Xinjiang, and we plan to extend the study to include more breeds in the future to address this issue. For example, we aim to study camels from different regions, environments, and breeds, thereby improving the practical applicability of the model developed in this research [37].

2) While we have achieved the expected results in addressing feature deformation issues through algorithm optimization, there is still a gap to higher performance metrics. Therefore, future work should explore methods to further improve the model's accuracy and performance.

Looking ahead, integrating more advanced computer vision techniques and incorporating multimodal frameworks could better facilitate information fusion [38], enhancing the model's ability to extract features. This would significantly improve the accuracy and practicality of pose estimation, especially for large-scale precision breeding farms with abundant resources. These improvements could make our framework more practical and promote the development of livestock farming in Xinjiang.

ACKNOWLEDGMENT

This research is funded by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (No. 2021D01C082). We would also like to express our sincere gratitude to Xinjiang Xintuo Dairy Group for their strong support.

REFERENCES

[1] Wu Saisai, Wu Jianzhai, Cheng Guodong, et al. Research progress on animal behavior recognition based on pose estimation in agricultural universities [J]. Journal of Agricultural University, 2023, 28(6): 22-35.

[2] Yang Liang, Wang Hui, Chen Ruipeng, et al. Research progress on pig-specific sensors [J]. Journal of Intelligent Agricultural Machinery (Chinese and English), 2023, 4(2): 22.

[3] He Dongjian, Liu Dong, Zhao Kaixuan. Research progress on intelligent perception and behavior detection of animals in precision livestock farming [J]. Journal of Agricultural Machinery, 2016, 47(5): 231-244.

[4] Yongfeng, Wang Wensheng, Guo Leifeng, et al. Research progress on daily behavior recognition of cattle based on wearable sensors [J]. Journal of Livestock Ecology, 2022, 43(10): 1-9.

[5] Guo Dongdong. Research on the recognition of goat behavior features based on triaxial acceleration sensor [D]. Taiyuan: Taiyuan University of Technology, 2015.

[6] Liu Haiying. Image processing technology based on computer vision algorithms [J]. Computer and Digital Engineering, 2019, 47(3): 672-677.

[7] Wang Hao. Application of computer vision in smart medical care [J]. Electronic Communication and Computer Science, 2024, 6(1): 86-88.

[8] Bai Yang, Bai Yun, Zhan Xini, et al. Technological characteristics and morphological evolution of artificial intelligence generated content (AIGC) [J]. Library and Information Science Knowledge, 2023, 40(1): 66-74.

[9] Yang Jiejing, Liu Zhigang. Exploration of the path of high-quality development of livestock farming under the model of **-style modernization [J]. Feed Research, 2023, 46(11).

[10] Yan Pengfei, Xiao Sha, Zhang Zhiyong, et al. Application and Research Progress of Computer Vision Technology in Modern Animal Husbandry[J]. **Pig Industry, 2024, 19(5): 83-89.

[11] Qi Yu, Su Han, Hou Rong, et al. A Method for Estimating Giant Panda Pose Based on High-Resolution Network[J]. Journal of Zoology in China, 2022, 42(4): 451.

[12] Chen Y, Fang R, Liang T, et al. Stock Price Forecast Based on CNN‐BiLSTM‐ECA Model[J]. Scientific Programming, 2021, 2021(1): 2446543.

[13] Fang H S, Xie S, Tai Y W, et al. Rmpe: Regional multi-person pose estimation[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2334-2343.

[14] Zheng C, Zhu X, Yang X, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and electronics in agriculture, 2018, 147: 51-63.

[15] Song H, Jiang B, Wu Q, et al. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline[J]. Transactions of the CSAE, 2018, 34(15): 190-199.

[16] Chen F, Liang X, Chen L, et al. Novel method for real-time detection and tracking of pig body and its different parts[J]. International Journal of Agricultural and Biological Engineering, 2020, 13(6): 144-149.

[17] Zheng W, Yan L, Chen L, et al. Knowledge-Embedded Mutual Guidance for Visual Reasoning[J]. IEEE Transactions on Cybernetics, 2023.

[18] Natesan B, Liu C M, Ta V D, et al. Advanced robotic system with keypoint extraction and YOLOv5 object detection algorithm for precise livestock monitoring[J]. Fishes, 2023, 8(10): 524.

[19] Rodriguez-Juan J, Berenguer-Agullo A, Benavent-Lledo M, et al. Bird Action Recognition in Wetlands using Deep Learning[C]//Proceedings of the 2024 International Conference on Information Technology for Social Good. 2024: 350-357.

[20] Research Progress and Technology Trend of Intelligent Monitoring of Dairy Cow Exercise Behavior[J]. Smart Agriculture, 2022, 4(2): 36.

[21] Wang J, Zhang X, Gao G, et al. Open Pose Mask R-CNN Network for Individual Cattle Recognition[J]. IEEE Access, 2023.

[22] Barney S, Dlay S, Crowe A, et al. Deep learning pose estimation for multi-cattle lameness detection[J]. Scientific Reports, 2023, 13(1): 4499.

[23] Mohanavel D, Ishwarya A M. Deep Learning and Computer Vision Based Warning System for Animal Disruption in Farming Environments[C]//2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT). IEEE, 2024: 1-6.

[24] Application of Deep Learning in Wildlife Conservation[J]. Acta Zoologica Sinica, 2023, 43(6): 734.

[25] Yuan, Zijian, et al. "YOLOv8-ACU: improved YOLOv8-pose for facial acupoint detection." Frontiers in Neurorobotics 18 (2024): 1355857.

[26] Deng, Ye, et al. "Context Adaptive Network for Image Inpainting." IEEE Transactions on Image Processing (2023).

[27] Wang, Shuqi, et al. "A Human Posture Estimation Method for Image Interaction System Based on ECA." International Forum on Digital TV and Wireless Multimedia Communications. Singapore: Springer Nature Singapore, 2023.

[28] Ren, Fei, et al. "Steel Surface Defect Detection Using Improved Deep Learning Algorithm: ECA-SimSPPF-SIoU-Yolov5." IEEE Access (2024).

[29] Peng, Yue, Alifu Kurban, and Mengmei Sang. "An Improved YOLOv8 Method for Measuring the Body Size of Xinjiang Bactrian

Camels." International Journal of Advanced Computer Science & Applications 15.8 (2024).

[30] Qu, X.; Long, Y.; Wang, X.; Hu, G.; Tao, X. Research on the Cable-to-Terminal Connection Recognition Based on the YOLOv8-Pose Estimation Model. Appl. Sci. 2024, 14, 8595.

[31] Ali Alameer. "Facial Emotion Recognition Datasets for YOLOv8Annotation." University of Salford, 2023.

[32] Hu Aijun, Tang Guiji, An Lianlian. Noise Reduction Method of Vibration Signal of Rotating Machinery Based on Mathematical Morphology[J]. Journal of Mechanical Engineering, 2006, 42(4): 127-130.

[33] Jiang T, Li Y, Feng H, et al. Research on a Trellis Grape Stem Recognition Method Based on YOLOv8n-GP[J]. Agriculture, 2024, 14(9): 1449.

[34] D. I. Krasnov, S. N. Yarishev, V. A. Ryzhova, and T. S. Djamiykov, "Improved YOLOv8 Network for Small Objects Detection," 2024 XXXIII International Scientific Conference Electronics (ET), 2024, doi: 10.1109/ET63133.2024.10721517.

[35] LI Menghe, XU Hongji, SHI Leixin, et al. Multi-person Behavior Recognition Based on Bone Key Point Detection[J]. Computer Science, 2021, 48(4): 138-143.

[36] Gong C, Zhang Y, Wei Y, et al. Multicow pose estimation based on keypoint extraction[J]. PloS one, 2022, 17(6): e0269259.

[37] Shao D, He Z, Fan H, et al. Detection of cattle key parts based on the improved Yolov5 algorithm[J]. Agriculture, 2023, 13(6): 1110.

[38] Yang Q, Liu Y, Lu L, et al. GFIDF: Gradual Fusion Intent Detection Framework[J]. 2024.