Preprints.org

Article

Multiple Postures and Behaviors Recognition Method for Horse Based on Se-SlowFast Network

Yanhong Liu , Jinxing Li , Fang Zhou , Xuan Chen , Wenxin Zheng , Tao Bai , Leifeng Guo [*] , Xinwen Chen [*]

Article

Multiple Postures and Behaviors Recognition Method for Horse Based on SE-SlowFast Network

Yanhong Liu [1,2,3,4], Jinxing Li [1,2,3,4], Fang Zhou [5], Xuan Chen [6], Wenxin Zheng [7], Tao Bai [1,3,4], Leifeng Guo [2,*] and Xinwen Chen [7,*]

[1] College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China; lzyfhf@gmail.com (Y.L.); lijinxing1129z@163.com (J.L.); bt@xjau.edu.cn(T.B.)

[2] Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100080, China

[3] Xinjiang Agricultural Informatization Engineering Technology Research Center, Urumqi 830052, China

[4] Ministry of Education Engineering Research Centre for Intelligent Agriculture, Urumqi 830052, China

[5] College of Information Science and Technology, Shihezi University, Shihezi 832000, China; zf_shzu@shzu.edu.cn (F.Z.)

[6] School of Software, Shanxi Agricultural University, Jinzhong 030801, China; 13223549193@163.com(X.C.)

[7] Institute of Animal Husbandry Quality Standards, Xinjiang Academy of Animal Science, Urumqi 830011, China; zwx2020@126.com (W.Z.)

* Correspondence: guoleifeng@caas.cn (L.G.); chxw106@163.com (X.C.)

Simple Summary: Horse behavior is a crucial indicator of equine health and welfare. Timely monitoring and analysis of various behaviors can provide valuable insights into the physiological state of horses. Unlike wearable devices, this study proposes an artificial intelligence algorithm to identify horse behaviors from video data. The method is designed to track and analyze multiple postures and behaviors, such as standing, crouching, lying down, feeding, and sleeping, in stabled horses. This approach can assist livestock managers in promptly assessing the health status of horses. Furthermore, this research offers new perspectives for equestrian clubs and modern horse breeding enterprises, contributing to the overall welfare of horses.

Abstract: The sleeping and grazing behaviors of horses play a crucial role in their health and welfare. In the horse industry, with the prevalent trend of individual single-stall farming in stable, accurately identifying various behaviors of horses within stables is imperative. This study has established a spatio-temporal dataset for horse postures and behaviors, proposed an algorithm utilizing a SE-SlowFast network with an enhanced loss function. This algorithm achieves automatic recognition of multiple postures and behaviors of horse simultaneously. The model developed in this study demonstrates high accuracy in identifying three postures within the stable: Standing, Crouching, and Lying, achieving respective accuracies of 92.73%, 91.87%, and 92.58%. Additionally, it accurately recognizes two behaviors: Sleeping and Grazing, achieving best accuracies of 93.56% and 98.77%. The overall best accuracy of the model is 93.9%. Experiments on video recognition confirm that the suggested model recognizes horse video data accurately. This study marks the first time computer vision technology has been used in the horse farming industry, achieving non-intrusive recognition of multiple horse postures and behaviors. This study provides valuable data support for livestock managers to evaluate the health conditions of horses, enhances the welfare of horses in stables, and contributes to the advancement of modern horse husbandry practices.

Keywords: horse; artificial intelligence; behaviors recognition; improved SlowFast network; animal welfare

---

1. Introduction

Horse is a common domestic animal, which is usually used for sports, companionship and various work roles [1]. At one time, horses were considered as important partners of human beings in agriculture and war, especially in transportation; Nowadays, horses play a major role in sports, leisure and tourism [2]. In the horse industry, especially for sports horses, continuous single stable

grazing (> 20h/day) is still dominant, and more and more scientific evidence shows that single stable grazing has a negative impact on the health of horses [3]. Horse postures and behaviors is one of the important factors to reflect its health status. When a horse has health problems, it will show different behavior characteristics. Monitoring the postures and behaviors of horses in stables and promptly assessing their health status is of paramount importance for modern horse industry management. However, in the face of continuous horses raised in a single stable, there are various problems such as high cost, high strength and easy fatigue in manual monitoring. Therefore, it is very necessary to use advanced equipment and technical means to quickly and efficiently identify horse postures and behaviors.

Recently, researchers have been exploring methods of animal behaviors recognition. Some researchers are committed to the research of intelligent wearable devices in animal behaviors recognition, and have achieved certain results. For example, using wearable equipment on sheep and using traditional machine learning algorithm, the classification of sheep behaviors [4,5] and the analysis of sheep behaviors when facing predators [6] are realized. Similarly, smart wearable devices are also applied to the behavior analysis of dairy cows, achieved the analysis of common behaviors of dairy cows [7,8] and the analysis of milk yield of dairy cows [9].

With the development of computer vision technology, researchers widely use deep learning to identify animal behaviors. Man et al. [10] used YOLO v5 to identify four behaviors of sheep, namely, lying, standing, grazing and drinking, and achieved an accuracy of 0.967. Yalei et al. [11] established a fusion network structure based on yolo and LSTM, and realized the identification of aggressive behavior of flock sheep, and achieved 93.38% accuracy. Hongke et al. [12] proposed a high-performance sheep image instance segmentation method based on Mask R-CNN framework, and the indexes of box AP, mask AP and boundary AP on the test set reached 89.1%, 91.3% and 79.5% respectively. Zishuo et al. [13] proposed a two-stage detection method based on yolo and VGG networks, which realized the identification of sheep behaviors and the classifications accuracy exceeded 94%. In the research field of cow behaviors recognition, various behavior recognition models based on yolo algorithm have also been widely used. For example, the recognition accuracy of routine behaviors such as drinking water, ruminating, walking, standing and lying down, grazing behavior and estrus behavior of dairy cows is higher than 94.3% [14–16]. Cheng et al. [17] proposed a method of cattle behavior recognition based on the dual attention mechanism, which combined the improved SE and CBAM attention modules with mobileNet network. The accuracy of this method can reach 95.17%. For other animal behaviors, such as pig and dog behavior recognition, researchers used convolutional neural network (CNN), long-term and short-term memory network (LSTM), DeepSORT network, yolox series network, ResNet50, PointNet and other networks to design an improved model based on the optimized parameters of the above networks or a multi-network fusion model to optimize the network structure. The recognition of routine behavior (walking, standing and lying down, etc.) [18,19], individual recognition [20], grazing time statistics [21], three-dimensional posture estimation [22,23], emotion recognition[24]and aggressive behavior recognition[25]are realized, and the accuracy rate is higher than 90%. Kaifeng et al. [26] proposed the video behavior recognition of pigs by combining the extended 3D convolution network (I3D) and the time sequence segmentation network (TSN), and the average accuracy rate of recognizing pigs' behaviors such as grazing, lying down, walking, scratching and crawling was 98.99%.

Conclusively, researchers have achieved excellent research results in their respective research fields by using different intelligent devices and advanced algorithms. However, in the field of intelligent recognition of horse behaviors, no related research has been found. This study fully consider and analyzed various research method put forward by researchers, and summarized that following views:

Smart wearable devices are contact data collection devices. Horses are more alert and sensitive than other domestic animals, especially when it comes to physical contact. As a result, there will be some interference with different horse postures and behaviors when using contact intelligent wearable devices. So, it is not the best idea to study horse postures and behaviors using contact-based intelligent wearable devices.

Compared with pictures, videos contain time series, and video behaviors recognition requires attention to both the target's spatial and temporal dimensions, making it richer in time domain information than picture behavior recognition alone. Therefore, video behaviors recognition has spatio-temporal characteristics. The research object of this study is the horse in stable, and its behaviors have very important spatio-temporal characteristics. Consequently, it is more meaningful to adopt horse video behavior recognition.

Behaviors recognition is a complex and challenging research. The common postures of horse include: Standing、Crouching and Lying.the common behaviors are: Sleeping and Grazing. Horses' behaviors all occur in a certain posture, so it is particularly important to apply the multi-label recognition algorithm to the recognition of horse postures and behaviors.

According to the above point of view, this study uses SlowFast [27] algorithm to recognize horse postures and behaviors. The algorithm can realize multi-tag recognition, and mark the result on the recognition object in the way of bounding box, which makes the result more intuitive and valuable. The contributions of this study are as follows:

(1). An AVA format dataset for horse multiple postures and behaviors recognition is constructed, which includes five categories: Standing, Crouching, Lying, Sleeping and Grazing.

(2). SE attention module is applied to the SlowFast network, and an improved loss function is proposed, which improves the accuracy of SlowFast in identifying horse posture and behavior.

(3). YOLOX is applied to the behaviors recognition of SlowFast network, which improves the efficiency of horse video recognition.

2. Materials and Methods

2.1. Description of Horse Postures and Behaviors

In the stable, most postures of horses are Standing, crouching and lying, and their behaviors are based on these three postures. Horses' sleeping behavior and grazing behavior are of great significance to their health. Horses that have no sleep due to prolonged activities will result in health detection and poor welfare, and made a detailed analysis with pictures of three sleeping positions [28]. Linda et al. [29] described the relationship between horse's sleep behavior and horse's health from four aspects, and put forward how to study horse's sleep behavior in the future. The lack of foraging opportunity leads to a reduction in caring time with consistent negative impacts on the digestive system and potential development of stereotypes [30]. In this study, computer vision algorithm will be used to study the sleeping behavior and grazing behavior of horses in a non-contact way, which will bring benefits to horse health. As shown in Figure 1, the original data supporting this study are all multi-angle and multi-directional video data resources from the experimental base.

(a) Stallion   (b) Mare and foal

Figure 1. Sample data of horse behaviors and postures: (a) is fine race horses; (b) is mares and a foals, foals lived with his mother until she was half a year old.

J´essica et al. [31] described the behavior of more than ten kinds of horses. The research team fully investigated the experimental base and put forward horses postures and behaviors suitable for this study. As shown in Table 1, there are seven kinds of tags for Multiple postures and behaviors

recognition of horse based on SE-SlowFast network, namely, Standing, Crouching, Lying, Standing-Sleeping, Standing-Grazing, Crouching-Sleeping and Lying-Sleeping.

Table 1. Definition of labels.

Label	Description
Standing	
Crouching	Standing、Crouching or Lying still, without entering a state of rest or engaging in sleep-related behaviors
Lying	
Standing, Grazing	Grazing in a standing
Standing, Sleeping	
Crouching, Sleeping	Sleeping in standing、crouching or lying
Lying, Sleeping	

2.2. Environmental Index

The postures and behaviors of horse are intricately linked with weather conditions, especially temperature [32]. Maintaining the horse's living environment within an optimal temperature range is crucial for conducting research on horse postures and behaviors. Therefore, in this study, intelligent sensors were installed on the beams of the stable to monitor environmental data such as temperature, illumination, and humidity, ensuring the scientific integrity of the research. As depicted in Figure 2, we records the key weather indices during the study period (from May 2023 to August 2023).
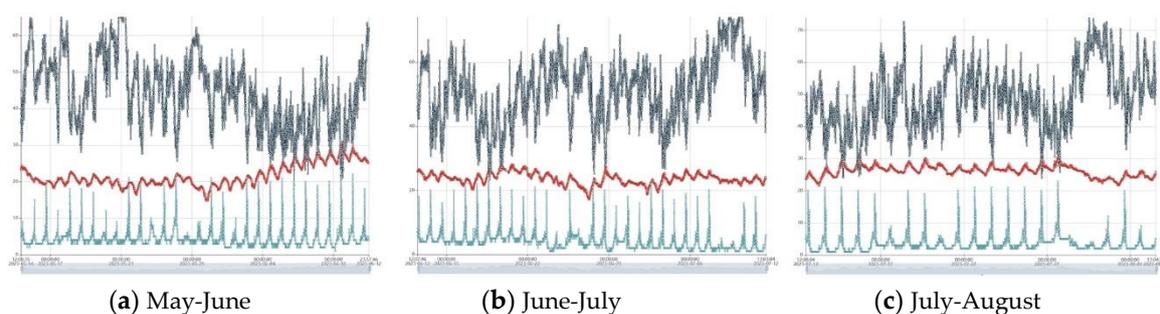

(**a**) May-June   (**b**) June-July   (**c**) July-August

Figure 2. Temperature, Illumination and Humidity Data in the stable: the average temperature(The red curve in the figure represents the temperature) is between 21.62 ℃ and 26.14 ℃, which is a suitable temperature. At the same time, the sensor can also collect ammonia, hydrogen sulfide and other environmental data in real time to ensure that the horses in the stable have a good living environment.

2.3. Data Acquisition

To efficiently and non-invasively collect horse postures and behaviors data without disturbing or affecting the horses' normal activities, this study employed smart cameras. As illustrated in Figure 3, a set of high-definition cameras with 360-degree rotation capability was strategically installed on the beams of the stable. The video streams were recorded at a resolution of 1920 x 1080 pixels with a frame rate of 30.0 frames per second. Each camera was positioned to cover a specific horse stall, comprehensively capturing the postures and behaviors of the horse in their respective stalls.
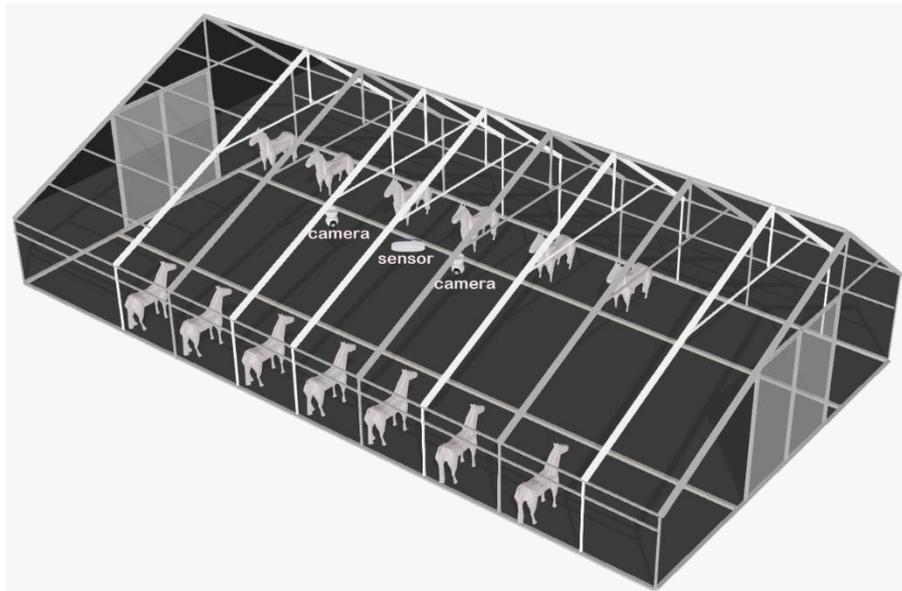
Figure 3. Schematic diagram of data collection scenario.

The data collection for this study was divided into two phases. The first phase spanned from May 2023 to August 2023, while the second phase took place from May to June 2024. Two sets of cameras were deployed: one for collecting data from stallions and the other for collecting data from mares and foals. A total of 48 horses, including stallions, mares, and foals, were captured in the video dataset. In total, 320 raw video clips were recorded, with a cumulative size of 9.08 TB. During the entire data collection process, continuous analysis of all the raw videos revealed a high degree of repetition in the postures and behaviors of horses within the stable. For most of the time, the horses in the stable were observed in a standing posture, occupying the majority of the 24-hour period. For example, the duration of sleep behavior in a standing posture was approximately 10 minutes, during which the horse remained nearly motionless. This posed challenges for subsequent video frame extraction. Additionally, due to the large body size of the horses, it was common for them to either turn their backs to the camera or extend their heads out of the stable, making it difficult to capture the postures and behavioral information accurately in the raw data. Based on these observations, this study systematically analyzed and organized all the posture and behavior videos collected during this period, forming a subset that represents the raw data of horse behavior for further analysis.

After excluding incomplete and highly repetitive data samples, this study retained 41 original videos capturing horse postures and behaviors, occupying a total space of 9.6 GB. As depicted in Figure 4, a subset of data samples utilized for horse behavior recognition research is showcased

Figure 4. Sample raw video data of horse postures and behaviors.

2.4. Dataset Construction

One of the focal points of this study is to construct two datasets: ODHS (Object Detection Dataset of Horse in Stable) for horse target detection and STHPB (Spatio-Temporal Dataset of Horse Postures and Behaviors) for spatio-temporal detection of horse postures and behaviors. As illustrated in Figure 5, ODHS is employed for detecting horses within the stable. STHPB is a spatio-temporal dataset focusing on horse postures and behaviors detection, encompassing all postures and behaviors data necessary for spatio-temporal detection of horses. The most important thing is that ODHS contains STHPB, which enables YOLOX to detect horses in all key frame pictures used in SE-SlowFast network when ODHS is applied to YOLOX.

(**a**) Sample data of ODHS(Object detection dataset of horse in stable)

b) Sample data of STHPB (Spatio-Temporal dataset of horse postures and behaviors)

Figure 5. ODHS and STHPB.

2.5. Model Implementation

To better recognize various horse postures and behaviors, we designed the identification framework as illustrated in Figure 6. In the task of recognizing multiple horse postures and behaviors, we replaced the original object detection algorithm, Faster R-CNN, integrated within the SlowFast network and instead adopted YOLOX for standalone object detection. This approach specifically identifies adult horses and foals in key frames. For posture and behavior recognition, we applied the improved SE-SlowFast algorithm. Finally, the RoIAlign method was employed to process the feature regions of interest for both object detection and behavior recognition, ensuring uniform sizing before performing classification predictions for various horse postures and behaviors.
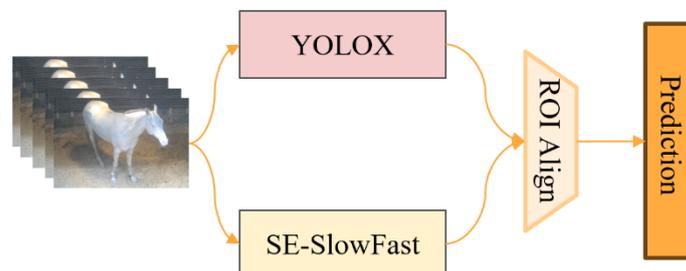

Figure 6. Architecture diagram for intelligent recognition of horse Multi-Posture and Behavior.

2.5.1. SE-SlowFast Network

The recognition of horse multiple postures and behaviors adopts a dual-stream network architecture known as the SE-SlowFast network. This network comprises three main components: Slow pathway with SE attention module added, fast pathway, and lateral connection, as depicted in Figure 7.
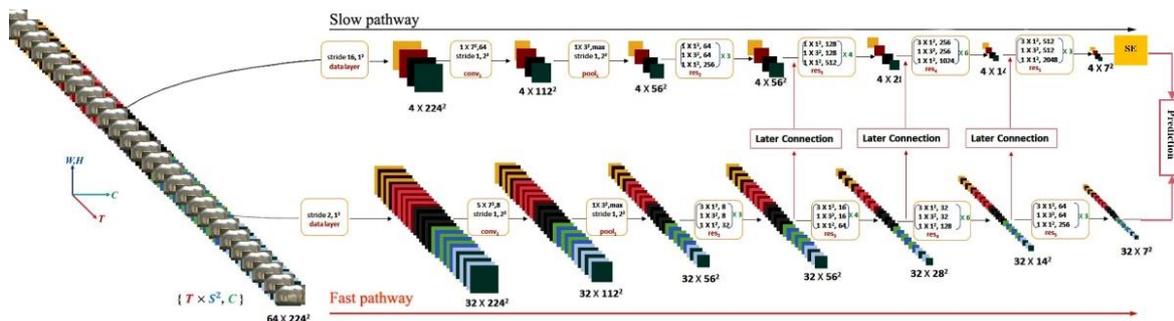
Figure 7. The architecture of spatiotemporal convolutional network for horse postures and behaviors recognition: The backbone network uses ResNet50, and the dimension size of the kernel is $\{T \times S^2, C\}$, where T represents the time dimension size, S^2 represents the spatial dimension size, and C represents the channel size.

2.5.1.1. SlowFast Network

The SlowFast network includes slow path and fast path. Slow Pathway: With an original clip length of 64 frames, the slow pathway's backbone network utilizes ResNet50. It has a larger time stride on input frames, set to 16 in this study, resulting in a refresh rate of approximately 2 frames per second, processing 1 frame out of 16. The number of frames sampled by the slow pathway is 4. Therefore, the slow pathway effectively addresses temporal down sampling. Fast Pathway: Compared to the slow pathway, the fast pathway possesses characteristics such as a high frame rate, high temporal resolution features, and low channel capacity. In terms of feature channels, the fast pathway uses a smaller time stride, set to 2 in this study. No temporal down sampling layer is used throughout the entire fast pathway until the global pooling layer before classification, ensuring the fast pathway maintains high-resolution features. Additionally, the fast pathway adopts a smaller number of channels, set to 1/8 of the slow pathway in this study. Given that the slow pathway has 64 channels, the fast pathway is configured with only 8 channels. This configuration ensures the fast pathway exhibits superior accuracy. After completing feature matching, the lateral connection links from the fast path to the slow path. Through multiple lateral connections, SlowFast achieves the fusion of feature information from both stream branches. Finally, the fused feature information is fed into the classifier for horse multifaceted behavior classification predictions.

2.5.1.2. SE Module

The basic structure of SE Module is shown in Figure 8. For any given transform F_{tr}: X→U, X∈R$^{W'×H'×C'}$, U∈R$^{W×H×C}$. The feature U first passes through F_{sq}, which can aggregate feature mapping across spatial dimension W×H to generate channel descriptors。This descriptor embeds the global distribution of channel characteristic response, so that the information from the global receptive field of the network can be used by its lower layers. After that, through $F_{sq.}$ the activation of specific samples is learned for each channel through the self-gating mechanism based on channel dependence, and the excitation of each channel is controlled. The feature map U is then re-weighted to generate the output of the SE block.
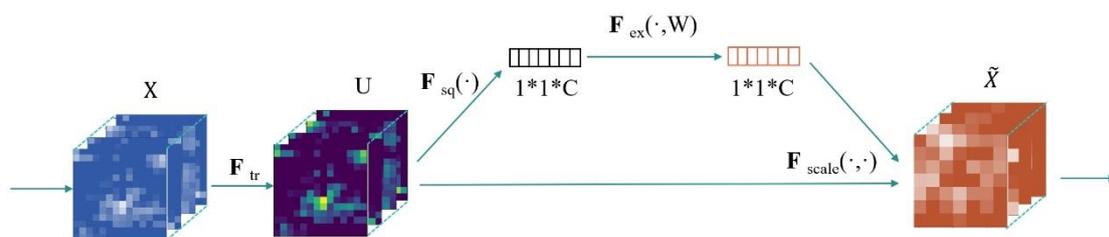

Figure 8. Structure diagram of SE Module.

2.5.2. YOLOX Network

To better identify adult horses and foals in key frames, as shown in Figure 9, we adopt YOLOX as the target detection algorithm, with yoloV3 and Darknet 53 as the baseline, and adopt the structure of Darknet 53 backbone network and SPP layer. The key structure of YOLOX is the use of Decoupled Head, which improves the convergence speed of YOLOX and realizes the end-to-end performance of YOLOX.
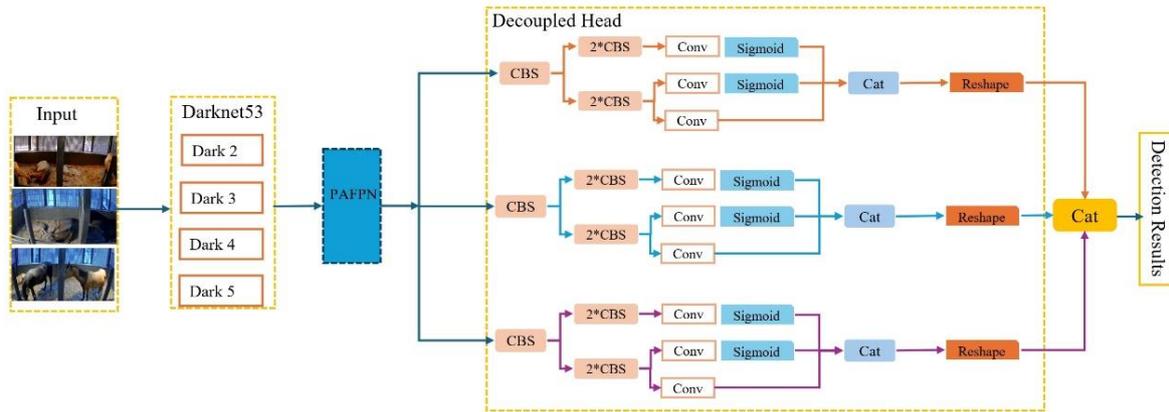
Figure 9. structural diagram of yolox.

2.6. Improved Loss Function

In classification tasks, the commonly used loss function for multilabel classification is Binary Cross Entropy Loss (BCE Loss). However, BCE Loss may not perform optimally when dealing with imbalanced datasets, leading to lower accuracy during model training. Therefore, this study addresses the characteristics of the STHPB dataset and adopts the CW Loss (Class Weighted Loss) combined with Focal Loss to form the CW_F_Combined Loss. This aims to reduce the weight of the loss function for categories with more instances and increase the weight for categories with fewer instances in the multilabel dataset. The formula for CW Loss is as follows:

$$L(y_i, \hat{y_i}, w_1) = -\frac{1}{N}\sum_{i=1}^{N}[w_i * (y_i * log(\sigma(\hat{y_i})) + (1 - y_i) * log(1 - \sigma(\hat{y_i})))] \tag{1}$$

$$W_i = \frac{\sum_{j=1}^{N} y_j}{N * y_{ji}}, for\ i\ =\ 1,2,\ldots,C \tag{2}$$

$$Focal\_Loss = r * (1 - p_t)^r * L, r\ =\ 2,3\ldots \tag{3}$$

Formula (1)is the definition of the CW Loss function, In the formulas,N represents the sample quantity,σ is the *sigmoid* function, $\hat{y_i}$ denotes the predicted score of the model,y_i signifies the true label, and w_i represents the class weight. Formula (2) is the definition of the class weight calculation formula, C is the total number of classes. The CW Loss consists of two main parts. CW Loss consists of two main components: the first part calculates the confidence of positive classes, and the second part calculates the confidence of negative classes. y_i=1, the first part predominates, maximizing the model's confidence in positive classes; y_i=0, the second part predominates, maximizing the model's confidence in negative classes. The purpose of w_i is to multiply the loss terms for each sample by the corresponding category weight during the computation of the loss function. This ensures that categories with fewer instances contribute more significantly to the loss during training.Formula (3) is the definition of *Focal_Loss*, Passing the value of $L(y_i, \hat{y_i}, w_1)$ to Focal_Loss, and the final loss function is computed, effectively addressing the issue of data imbalance.

2.7. Data Enhancement

The horse raw video data exhibits a severe imbalance issue, with the majority of data representing standing postures, while crouching and lying postures have fewer instances. Despite proposing the CW_F_Combined Loss in this study to address this phenomenon, the significant data imbalance still impacts the overall performance of the model. To mitigate the impact on the overall model performance while maintaining effective horse postures and behaviors video recognition, this study aims to integrate and reduce the sample count of standing postures as much as possible. video

data augmentation and temporal frame image data augmentation techniques are employed to augment the sample count of crouching and lying postures.

2.7.1. Video Data Enhancement

In this study, video processing software, as shown in Figure 10, is utilized to crop and concatenate all original crouching and lying posture video data, enabling horizontal flipping to expand the sample count of these postures.
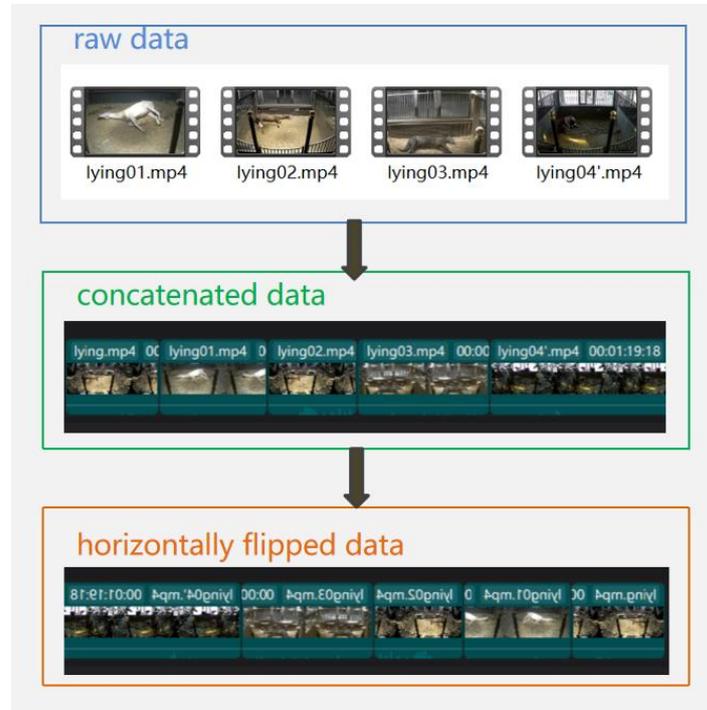

Figure 10. Samples of horse raw video data augmentation：The study only used horizontal flip to process video data, did not use other methods to expand video data, such as 90-degree rotation, 180-degree rotation and other angle rotations.

2.7.2. Image Data Enhancement

After augmenting the horse postures and behaviors video data through video data augmentation, this study employs image data augmentation techniques to further enhance the dataset for crouching and lying postures. To ensure optimal model performance, as illustrated in Figure 11, this study selectively employs three methods for image augmentation: Color Jittering, Adding Noise, and CLAHE. The augmentation process is randomized for variability.
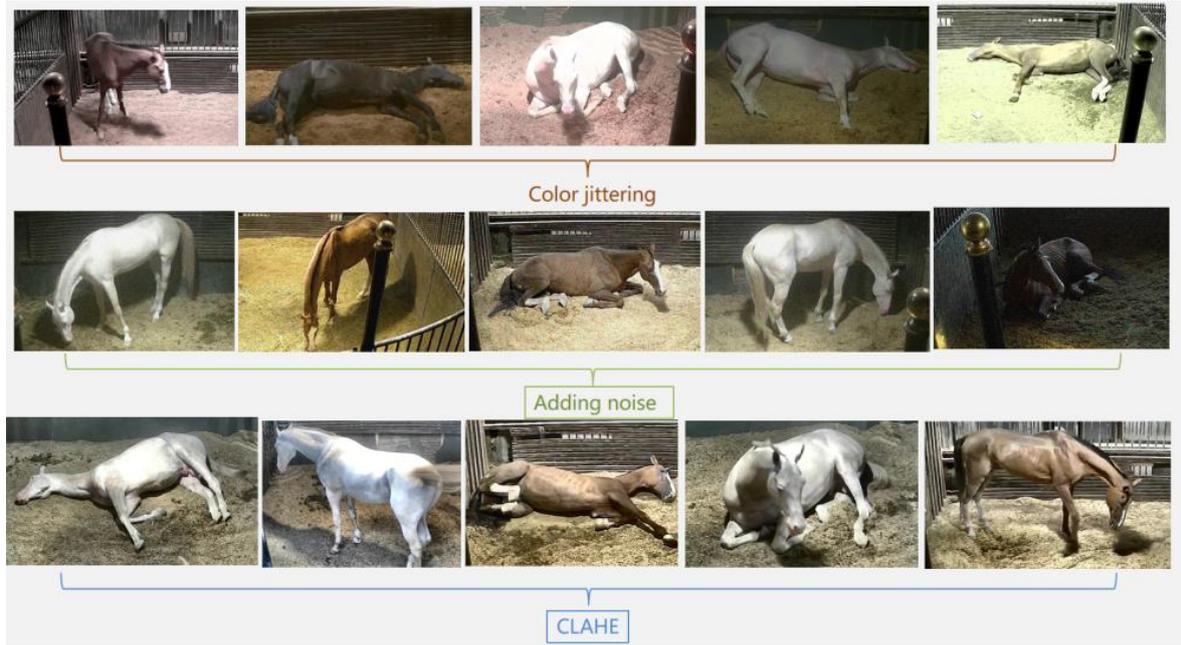
Figure 11. Example of Image data augmentation.

2.8. Model Evaluation Metrics

To objectively analyze the model's performance, this study adopts the Mean Average Precision (mAP) with IOU=0.5 as an evaluation metric. The formula is as follows:

$$IOU(A, B) = \left|\frac{A \cap B}{A \cup B}\right| \tag{4}$$

$$mAP = \int_0^1 P(R)dR \tag{5}$$

$$R = \frac{T_{rue}P_{ositive}}{T_{rue}P_{ositive} + F_{alse}N_{egative}} \tag{6}$$

$$P = \frac{T_{rue}P_{ositive}}{T_{rue}P_{ositive} + F_{alse}P_{ositive}} \tag{7}$$

Formula (4) represents the Intersection Over Union (IOU) threshold calculation, where A denotes the ground truth bounding box, and B represents the predicted detection box. Formula (5) is the calculation formula for mAP, where R is recall, and P is precision. Formulas (6) and (7) provide the calculation formulas for R and P, $T_{rue}P_{ositive}$ representing the number of true positives; $F_{alse}N_{egative}$ representing the number of false positives; $F_{alse}P_{ositive}$ represents the number of negative samples mistakenly classified as positive samples; represents the total number of positive samples; $T_{rue}P_{ositive} + F_{alse}N_{egative}$ represents the total number of positive samples; $T_{rue}P_{ositive} + F_{alse}P_{ositive}$ represents the total number of samples classified as positive.

3. Experimentation and Results

3.1. Experiment Implementation Details

Figure 12 illustrates the primary workflow of the conducted experiments. The research team deployed high-definition cameras at elevated positions along the corridor of the stable. The captured video data was stored on a cloud server, followed by downloading and data preprocessing. Prior to model training, the preprocessed video data underwent further processing to create the COCO dataset and AVA dataset. The YOLOX network and SE-SlowFast network were then employed to separately train models, resulting in the horse object recognition model and the horse Multiple postures and behaviors recognition model. Finally, the original horse video data underwent

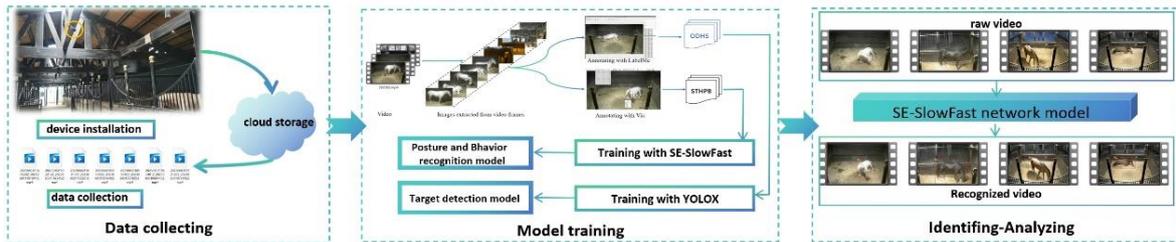recognition and interpretation, and the recognition results were systematically analyzed and documented.

Figure 12. Process of experiment implementation.

3.2. Performance Evaluation

3.2.1. Feature Learning of SE-SlowFast Network

To verify the effect of SE-SlowFast on multi-pose and behavior recognition of horses. We visualizes the output features of the Slow pathway, responsible for postures and behaviors feature extraction in the SE-SlowFast network. The four layers visualized are before the lateral connections, corresponding to stages 2, 3, 4, and 5. As depicted in Figure 13, after Res2, the Slow pathway has already learned features from the input images, and as the network deepens, the feature learning becomes more profound.

Figure 13. Example of Slow pathway Feature Learning：Res2,Res3,Res4,Res5correspond to Figure 6. Each feature map learned after the convolution operation has sizes: $56^2, 28^2, 14^2, 7^2$.

3.2.2. Loss Function Comparison

This study compared the training performance of BCE_F_Combined Loss (Binary CrossEntropy Loss combined with Focal Loss) and CW_F_Combined Loss (Class Weighted Loss combined with Focal Loss). A total of 6 rounds of experiments were conducted, with different values assigned to r for Focal_Loss in each training round. Figure 14 presents the training loss curves for the SlowFast network on the horse postures and behaviors dataset during continuous iterations for both BCE_F_Combined Loss and CW_F_Combined Loss,the convergence speed was fast and the oscillation was smooth before reaching convergence.
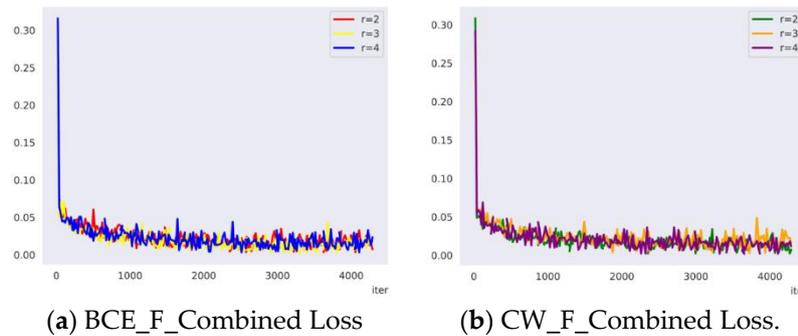

(**a**) BCE_F_Combined Loss   (**b**) CW_F_Combined Loss.

Figure 14. Comparison of two different loss functions in the SE-SlowFast network.

As shown in Table 2, BCE_F_Combined Loss achieved the expected results for predicting certain postures or behaviors, performing even better in some cases. For instance, when mAP is optimized, the accuracy of Sleeping and Grazing is 92.98% and 98.22%, respectively. However, the accuracy of Crouching and Lying is both below 90%, indicating that the overall performance of the model is not sufficiently excellent. In comparison to BCE_F_Combined Loss, CW_F_Combined Loss accurately calculates the weights in Formula (2) for this research dataset. The experiments demonstrate that CW_F_Combined Loss performs more outstandingly on the STHPB. When r=2, compared to BCE_F_Combined Loss, all categories of postures and behaviors have seen improvements. Specifically, the accuracy of Crouching and Lying, which have a lower proportion in the STHPB, increased by 3.44% and 3.01%, respectively. The overall mAP is 93.90%, an increase of 1.57%.

Table 2. Accuracy comparison of SE-SlowFast training with different loss functions.

Loss Functions		postures			Behaviors		mAP
		Standing	Crouching	Lying	Sleeping	Grazing	(IOU=0.5)
BCE_F_Combined Loss	r=2	0.9126	0.8724	0.9145	0.9188	0.9621	0.9161
	r=3	0.9121	0.8915	0.9089	0.9051	0.9745	0.9182
	r=4	0.9245	0.8843	0.8957	0.9298	0.9822	0.9233
CW_F_Combined Loss	r=2	0.9273	0.9187	0.9258	0.9356	0.9877	0.9390
	r=3	0.9156	0.8834	0.9102	0.9213	0.9544	0.9170
	r=4	0.9135	0.8525	0.9012	0.9124	0.9725	0.9104

3.2.3. Ablation Study

To evaluate the role of SE module in SlowFast and visually demonstrate the improvement of the enhanced SlowFast model in detecting various behaviors, we conducted ablation experiments. The detection results of multiple postures and behaviors of horses are compared. Table 3 shows the results of ablation experiments. √ indicates that the module is used, and × indicates that the module is not used. When the SlowFast network does not introduce the SE attention module, the overall accuracy is low, and the highest accuracy is Grazing, which is 92.47. After adding SE attention module to the front end of Slow Path, the overall accuracy is not improved obviously, and some accuracy has been improved, such as the accuracy of Standing, Crouching and Sleeping increased by 0.78%, 0.71% and

1.12% respectively, but the accuracy of Lying and Grazing decreased. When the SE attention module is added to the back end of the Slow Path, the overall performance of the SlowFast network is obviously improved, and both the accuracy of logistics and behaviors are improved, especially the accuracy of Grazing reaches 98.77%. Therefore, the use of SE attention module in SlowFast network is beneficial to identify various postures and behaviors of horses.

Table 3. Results of ablation experiments.

Backbone Network	SE Module at the Front of the Slow Path	SE Module at the End of the Slow Path	Accuracy of postures and Behaviors Recognition				
			Standing	Crouching	Lying	Sleeping	Grazing
	×	×	0.8945	0.9051	0.8829	0.9011	0.9247
SlowFast	√	×	0.9023	0.9122	0.8754	0.9123	0.9189
	×	√	0.9273	0.9187	0.9258	0.9356	0.9877

3.3. Recognition Effect

SlowFast relies on target detection algorithm for behavior recognition, and YOLOX is used as the detector in this study. As shown in Table 3. The experiment evaluates the speed of four detection algorithms by setting the bounding box obtained by IOU to 0.5. In Table 4, this section takes 10 videos as experiments and takes the detection time as an indicator. The length of each video is 60 seconds. Firstly, the horse in the video is identified, and then the spatiotemporal behavior of the horse is analyzed. Combining the horse detection time and spatio-temporal motion detection time in Table 5, compared with FasterCNN+SE-SlowFast, YOLOv3+SE-SlowFast and YOLOX+SE-SlowFast significantly improve the processing time of the algorithm, and YOLOX+SE-SlowFast has the best processing speed and faster processing speed. Therefore, YOLOX can be used as the horse multi-pose and behavior target detection algorithm to realize faster recognition of horses.

Table 4. Comparison of different algorithms for video frame detection and Spatio-Temporal Action Detection time.

Method	Horse Detection Time	Spatio-Temporal Action Detection Time
FasterRCNN+ SE-SlowFast	35s	12.5s
YOLOv3 + SE-SlowFast	13s	11s
YOLOX + SE-SlowFast	12.5s	10s

Also, We selected several representative videos for recognition test to evaluate the performance of the model. As shown in Figure 15**.**, the experimental results show that the SE-SlowFast model performs relatively well in each video, and can accurately identify Standing, Crouching, Lying, Sleeping and Grazing. The multi-pose and behavior video recognition of horses is realized.
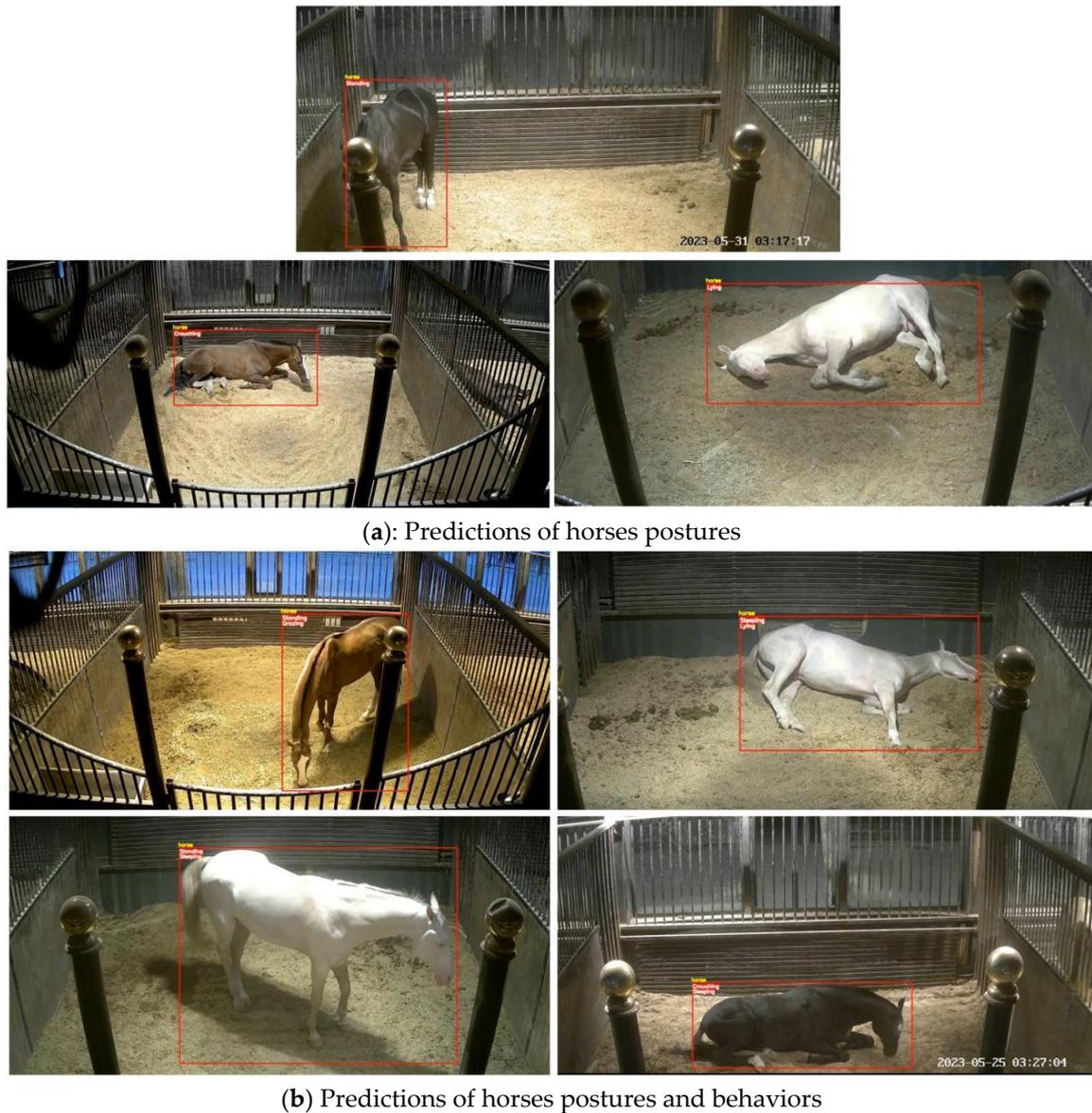
(**a**): Predictions of horses postures

(**b**) Predictions of horses postures and behaviors

Figure 15. Examples of predicting horse multi-postures and behaviors using the optimal model.

The research team recorded detailed recognition data using the following method: recognition results were recorded every 1 second, with a total of 60 recordings per video. As depicted in Figure 16, false positives occurred in cases (a), (d), and (g) with rates of 6%, 6.8%, and 6.8%; (c) and (f) showed missed detections, with rates of 1.6% and 3.3%. On the other hand, (b), (e), (h), and (i) demonstrated good recognition performance. The experiments confirm that the horse multi-postures and behaviors recognition model based on the SE-SlowFast network exhibits overall good performance, with only a few instances of false positives and missed detections.
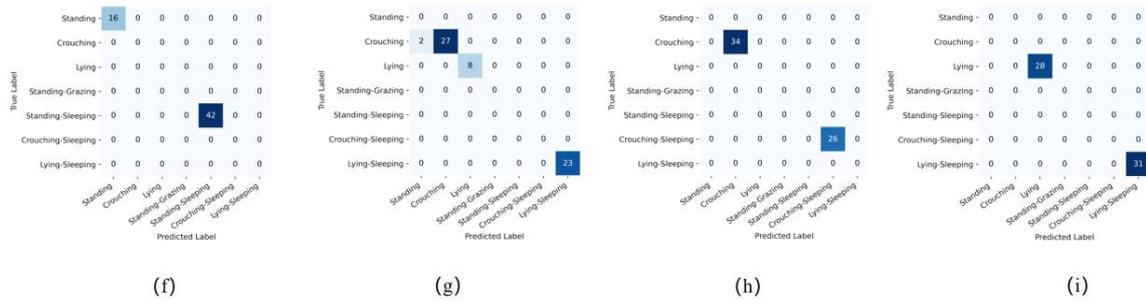
(f)　　　　　　　　(g)　　　　　　　　(h)　　　　　　　　(i)
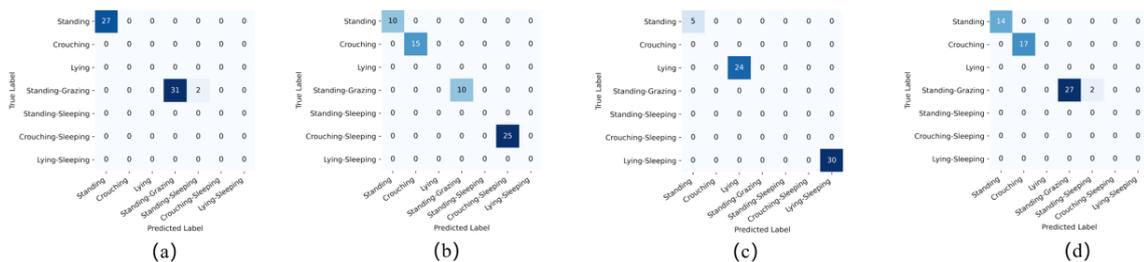
Figure 16. Confusion Matrix for Video Recognition.

4. Discussion

4.1. The Design of the Recognition System

The overall size of the framework designed for horse multi-postures and behaviors in this study is 557MB, presenting a challenge for the deployment of the algorithm, requiring computer support with GPU capabilities. Therefore, in the initial stages of the research, there is a plan to deploy the entire model using a cloud platform to analyze the captured video from cameras. As shown in Figure 17, the research team has developed the "Real-time Recognition System for Horse Multi-Postures and Behaviors." This system can recognize horse video data captured by cameras and push the recognition results to the "Horse Smart Farming Big Data Platform", designed by the research team in 2023 [33]. This not only benefits the health and welfare of the horse but also provides convenient online access for livestock managers to monitor the conditions of horses in the stable.

Figure 17. Real-time Recognition System for Horse Multi-Postures and Behaviors.

4.2. Comprehensive Deployment of Intelligent Camera Devices

Despite addressing concerns related to the positioning of horses within the stable, some limitations persist. Specifically, difficulties arise when the horse's back is facing the camera, making it challenging to capture the horse's head and subsequently hindering accurate behavior assessment. This situation amplifies the rates of false positives and missed detections in video recognition. Therefore, in the next phase of the research, we plan to install intelligent camera devices at diagonal

positions within the stable to comprehensively monitor various postures and behaviors of the horse, aiming to mitigate the aforementioned challenges.

4.3. Algorithm Optimization

The SlowFast algorithm utilizes ResNet50 as its backbone, which is a deep convolutional neural network with 50 layers. Given that the study focuses on horse within stables, scenarios where multiple horses appear within the camera's range are uncommon. The collected raw video data typically features a limited number of horses in the same scene. In future research, We will conduct a more in-depth study on the SlowFast algorithm, exploring the use of lightweight neural network architectures for algorithm optimization to enhance model applicability.

5. Conclusions

This study established a spatio-temporal horse postures and behaviors dataset, a SE-SlowFast network to automatically recognize multiple posterior and behaviors of horse is proposed. This study, utilizing computer vision technology, achieves the first-ever recognition of multiple horse postures and behaviors, contributing to the welfare and health of horse. Experimental results indicate that the trained model in this study achieves a recognition accuracy of 93.9%. Compared to traditional single-label recognition techniques, this study realizes simultaneous multi-label recognition, providing richer identification information. This approach offers more effective technical support for non-contact monitoring of horse postures and behaviors in stables. Recently, the research team has been actively involved in the development of key technologies for managing active horses and intelligent equestrian facilities. We have successfully applied big data technology and computer vision techniques to modernize horse farming practices.

In the near future, while adhering to the principle of non-intrusive and non-disruptive monitoring of horses' normal lives, the research team will continue to delve into the application of advanced artificial intelligence technology in various aspects of horse production and living. This ongoing research aims to further enhance the welfare of horses and contribute to their overall health.

Author Contributions: Yanhong Liu: Data curation, Formal analysis, Software, Visualization, Writing-original draft. Jinxing Li: Software, Writing – review & editing. Fang Zhou: Software. Xuan Chen: Data curation, Resources. WenXin Zheng: Resources, Writing-review & editing. Tao Bai: Investigation, Validation. Leifeng GUO: Conceptualization, Methodology, Resources, Writing-review & editing. Xinwen Chen: Methodology, Writing – review & editing. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: All procedures were approved by the Xinjiang Institute of Animal Husbandry and Veterinary Sciences for experimental animal welfare and ethics.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: Data are available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Brubaker, L.; Udell, M.A.R. Cognition and Learning in Horses (Equus caballus): What We Know and Why We Should Ask More. *Behav. Process.* **2016**, 126, 121–131. [CrossRef]
2. Danby, P.; Grajfoner, D. Human–Horse Tourism and Nature-Based Solutions: Exploring Psychological Well-Being Through Transformational Experiences. *J. Hosp. Tour. Res.* **2022**, 46, 607–629. [CrossRef]

3. Lesimple, C.; Reverchon-Billot, L.; Galloux, P.; Stomp, M.; Boichot, L.; Coste, C.; Henry, S.; Hausberger, M. Free Movement: A Key for Welfare Improvement in Sport Horses? *Appl. Anim. Behav. Sci.* **2020**, 225, 104972. [CrossRef]

4. Fogarty, E.S.; Swain, D.L.; Cronin, G.M.; Moraes, L.E.; Trotter, M. Behaviour Classification of Extensively Grazed Sheep Using Machine Learning. *Comput. Electron. Agric.* **2020**, 169, 105175. [CrossRef]

5. Price, E.; Langford, J.; Fawcett, T.W.; Wilson, A.J.; Croft, D.P. Classifying the Posture and Activity of Ewes and Lambs Using Accelerometers and Machine Learning on a Commercial Flock. *Appl. Anim. Behav. Sci.* **2022**, 251, 105630. [CrossRef]

6. Evans, C.A.; Trotter, M.G.; Manning, J.K. Sensor-Based Detection of Predator Influence on Livestock: A Case Study Exploring the Impacts of Wild Dogs (Canis familiaris) on Rangeland Sheep. *Animals* **2022**, 12, 219. [CrossRef]

7. Wu, Y.; Liu, M.; Peng, Z.; Liu, M.; Wang, M.; Peng, Y. Recognising Cattle Behaviour with Deep Residual Bidirectional LSTM Model Using a Wearable Movement Monitoring Collar. *Agriculture* **2022**, 12, 1237. [CrossRef]

8. Balasso, P.; Taccioli, C.; Serva, L.; Magrin, L.; Andrighetto, I.; Marchesini, G. Uncovering Patterns in Dairy Cow Behaviour: A Deep Learning Approach with Tri-Axial Accelerometer Data. *Animals* **2023**, 13, 1886. [CrossRef]

9. Scheurwater, J.; Jorritsma, R.; Nielen, M.; Heesterbeek, H.; van den Broek, J.; Aardema, H. The Effects of Cow Introductions on Milk Production and Behaviour of the Herd Measured with Sensors. *J. Dairy Res.* **2022**, 88, 374–380. [CrossRef]

10. Cheng, M.; Yuan, H.; Wang, Q.; Cai, Z.; Liu, Y.; Zhang, Y. Application of Deep Learning in Sheep Behavior Recognition and Influence Analysis of Training Data Characteristics on the Recognition Effect. *Comput. Electron. Agric.* **2022**, 198, 107010. [CrossRef]

11. Xu, Y.; Nie, J.; Cen, H.; Wen, B.; Liu, S.; Li, J.; Ge, J.; Yu, L.; Pu, Y.; Song, K.; Liu, Z.; Cai, Q. Spatio-Temporal-Based Identification of Aggressive Behavior in Group Sheep. *Animals* **2023**, 13, 2636. [CrossRef]

12. Zhao, H.; Mao, R.; Li, M.; Li, B.; Wang, M. SheepInst: A High-Performance Instance Segmentation of Sheep Images Based on Deep Learning. *Animals* **2023**, 13, 1338. [CrossRef]

13. Gu, Z.; Zhang, H.; He, Z.; Niu, K. A Two-Stage Recognition Method Based on Deep Learning for Sheep Behavior. *Comput. Electron. Agric.* **2023**, 212, 108143. [CrossRef]

14. Qiang, B.; Ronghua, G.; Chunjiang, Z. Multi-Scale Behavior Recognition Method for Dairy Cows Based on Improved YOLOV5s Network. *Trans. Chin. Soc. Agric. Eng.* **2022**, 38, 163–172. [CrossRef]

15. Wang, R.; Gao, Z.; Li, Q.; Zhao, C.; Gao, R.; Zhang, H.; Li, S.; Feng, L. Detection Method of Cow Estrus Behavior in Natural Scenes Based on Improved YOLOv5. *Agriculture* **2022**, 12, 1339. [CrossRef]

16. Yu, Z.; Liu, Y.; Yu, S.; Wang, R.; Song, Z.; Yan, Y.; Li, F.; Wang, Z.; Tian, F. Automatic Detection Method of Dairy Cow Grazing Behaviour Based on YOLO Improved Model and Edge Computing. *Sensors* **2022**, 22, 3271. [CrossRef]

17. Shang, C.; Wu, F.; Wang, M.; Gao, Q. Cattle Behavior Recognition Based on Feature Fusion Under a Dual Attention Mechanism. *J. Vis. Commun. Image R.* **2022**, 85, 103524. [CrossRef]

18. Tu, S.; Zeng, Q.; Liang, Y.; Liu, X.; Huang, L.; Weng, S.; Huang, Q. Automated Behavior Recognition and Tracking of Group-Housed Pigs with an Improved DeepSORT Method. *Agriculture* **2022**, 12, 1907. [CrossRef]

19. Kim, J.; Moon, N. Dog Behavior Recognition Based on Multimodal Data from a Camera and Wearable Device. *Appl. Sci.* **2022**, 12, 3199. [CrossRef]

20. Zhou, H.; Li, Q.; Xie, Q. Individual Pig Identification Using Back Surface Point Clouds in 3D Vision. *Sensors* **2023**, 23, 5156.[CrossRef]

21. Chen, C.; Zhu, W.; Steibel, J.; Siegford, J.; Han, J.; Norton, T. Recognition of Grazing Behaviour of Pigs and Determination of Grazing Time of Each Pig by a Video-Based Deep Learning Method. *Comput. Electron. Agric.* **2020**, 176, 105642. [CrossRef]

22. Yu, R.; Choi, Y. OkeyDoggy3D: A Mobile Application for Recognizing Stress-Related Behaviors in Companion Dogs Based on Three-Dimensional Pose Estimation through Deep Learning. *Appl. Sci.* **2022**, 12, 8057.[CrossRef]

23. Ji, H.; Yu, J.; Lao, F.; Zhuang, Y.; Wen, Y.; Teng, G. Automatic Position Detection and Posture Recognition of Grouped Pigs Based on Deep Learning. *Agriculture* **2022**, 12, 1314. [CrossRef]

24. Chen, H.-Y.; Lin, C.-H.; Lai, J.-W.; Chan, Y.-K. Convolutional Neural Network-Based Automated System for Dog Tracking and Emotion Recognition in Video Surveillance. *Appl. Sci.* **2023**, 13, 4596. [CrossRef]

25. Ji, H.; Teng, G.; Yu, J.; Wen, Y.; Deng, H.; Zhuang, Y. Efficient Aggressive Behavior Recognition of Pigs Based on Temporal Shift Module. *Animals* **2023**, 13, 2078.[CrossRef]

26. Zhang, K.; Li, D.; Huang, J.; Chen, Y. Automated Video Behavior Recognition of Pigs Using Two-Stream Convolutional Networks. *Sensors* **2020**, 20, 1085. [CrossRef]

27. Feichtenhofer, C.; Fan, H.; Malik, J.; He, K. SlowFast Networks for Video Recognition. In Proceedings of the IEEE International Conference on Computer Vision, **2019**. [CrossRef]

28. Lim Teik Chung, E.; Khairuddin, N.H.; Tengku Azizan, T.R.P.; Adamu, L. Sleeping Patterns of Horses in Selected Local Horse Stables in Malaysia. *J. Vet. Behav.* **2018**, 26, 1-4. [CrossRef]

29. Greening, L.; McBride, S. A Review of Horse Sleep: Implications for Horse Welfare. *Front. Vet. Sci.* **2022**, 9, 916737. [CrossRef]

30. Ellis, A.D.; Fell, M.; Luck, K.; Gill, L.; Owen, H.; Briars, H.; Barfoot, C.; Harris, P. Effect of Forage Presentation on Feed Intake Behaviour in Stabled Horses. *Appl. Anim. Behav. Sci.* **2015**, 165, 88–94.[CrossRef]

31. Seabra, J.C.; do Vale, M.M.; Spercoski, K.M.; Hess, T.; de Moura, P.P.V.; Dittrich, J.R. Time-Budget and Welfare Indicators of Stabled Horses in Three Different Stall Architectures: A Cross-Sectional Study. *J. Equine Vet. Sci.* **2023**, 131, 104936.[CrossRef]

32. Aviron, M.; Geers, R. Behavior of Horses on Pasture in Relation to Weather and Shelter: A Field Study in a Temperate Climate. *J. Vet. Behav.* **2015**, 10, 561-568.[CrossRef]

33. Liu, Y.; Cao, K.; Chen, X.; Li, J.; Xiong, T.; Du, X.; Bai, T.; Zheng, W.; Guo, L. Design and Implementation of a Horse Intelligent Breeding Big Data Platform. *J. Agric. Big Data* **2023**, 5, 93–103. [CrossRef]