# A Two-stage Sheep Acial Pain Recognition Method Based on Deep Learning

LI Huan, HE Lile, HE Ning, ZHANG Chenrui

(*College of Electrical and Mechanical Engineering*, *Xi'an University of Architecture and Technology*, *Xi'an*, *710055*, *China*)

**Abstract:** In order to address the issues of the current sheep face pain detection algorithm under complex environments with poor detection accuracy and complex models, this paper proposes a two-stage method of sheep face pain detection based on light yolov5s. First, the weights of the yolov5s model are reduced by combining the Ghostnet structure, feature fusion is performed using the BiFPN structure and the ODConv join to improve detection accuracy. The experimental results show that the number of parameters and complexity of the optimized model are reduced by 38.03% and 50.94%, respectively, compared with the original model, and the accuracy of recognizing sheep faces is improved by 0.2% and the recall rate is increased by 1%. Compared to current mainstream algorithms such as yolov4 tiny and SSD, it not only significantly reduces the number of parameters but also has a better detection performance. Second, pain detection of sheep faces recognised by lightweight yolov5s was carried out by MobileNetV2, and experimental results showed that MobileNetV2 achieved a mean classification accuracy of over 98% for painful sheep, and this illustrates that this two-stage sheep face pain recognition research method has great application value for sheep health breeding.

**Keywords:** Pain Detection; Yolov5 Jightweight; BiFPN; ODConv

## 1 Introduction

Animal welfare and healthy growth are crucial for the agricultural industry. Failure to treat animals in a timely manner after illness can cause group infection or animal death in a short period of time, resulting in huge losses to the farm. Many animals are often sick with signs of pain, and the lack of recognition of those signs will result in failure to treat animals in a timely fashion, and is one of the most commonly cited reasons for production impacts[1]. Facial expressions have been used as an indicator of pain levels for a variety of animals[2]-[4]. Identification and quantification of pain levels are necessary for subsequent treatment and pain relief[5], and the presence of painful conditions in sheep is often indicative of diseases such as sheep rotting hoof disease[6], mastitis[7], and pregnancy toxaemia[8]. Literature[9] created the Sheep Pain Facial Expression Scale (SPFES), which determined pain levels based on orbital tightness, cheek tightness, ear position, lip and jaw outlines, as well as the position of the nostrils and nasal grooves, and achieved high accuracy in the identification of pain and disease prediction for sheep. However, this method required professional training of breeders and can be highly subjective. These scoring standards using manual detection suffered from inefficiency and poor recognition accuracy. At present, the application of computer vision technology in the farming industry has performed well in the areas of face recognition, disease prediction and abnormal behaviour detection[10]-[13]. In paper[14]-[15], pain detection in horses and sheep was performed automatically using computer vision techniques ,and more specifically, those methods were performed by first extracting the region of interest, performing region-based feature extraction, and then classifying the pain based on the gradient histogram (HOG), or a facial action unit (AU) to assess the level of pain. The paper[16] employed computer vision techniques and further proposed a method of transfer learning and fine-tuning that is capable to detect the facial pain in sheep with complex backgrounds.

However, in the actual process of sheep captivity, shading, background complexity, and other issues will manifest. These factors will interfere with the identification results, and lead to identification errors that may cause the farm management order to chaos. In addition, these algorithms are redundant, difficult to implement with lightweight inspection equipment, and have certain limitations. In recent years, numerous scholars have conducted in-depth research on lightweight detection models. Yu L[17] replaced the original Darknet-53 with EfficientNet-B0 for feature extraction in YOLO V3 neural network, thereby making the model lighter and easier to deploy and thereby decreasing detection time. Song S[18] enhanced yolov3 based on the K-mean clustering algorithm with a model compression method combining channel pruning and layer pruning, and applied it to the identification of individual sheep. Additionally, the size of the improved model is reduced to one-quarter of the original one. Li X[19] combined Mobilenetv2 with a Vision Transformer to improve the model's ability to extract fine-grained features. In comparison to models like Swin-small, which currently performs SOTA, the parameters and FLOPs were reduced by nearly a factor of ten, while the recognition accuracy was reduced by only 0.64 percent. However, relatively little research has been conducted on the application of lightweight model in actual farming environments, and further research is required to determine how to apply lightweight network models to sheep face pain detection. This paper proposes a two-stage sheep face pain detection method based on lightweight yolov5s based on the above comparative analysis of different lightweight methods and the complex situation of farm background. In addition to lightened the model, this method also avoided the interference of the complex backgrounds on the recognition results.

## 2 Materials and Methods

### 2.1 Experimental Environment

The experimental environment is configured on Windows 10 operating system, NVIDIA GeForce RTX 3080, cuda11.0, python 3.7, and pytorch1.2.0 framework is used for training. The training parameters of each model are shown in Table 1.

**Table 1 Model Training Parameters**

| Training Parameters | Lightened Yolov5 | MobileNetV2 |
|---|---|---|
| Optimization Algorithms | SGD | SGD |
| Image Dimensions | 640×640 | 224×224 |
| Learning Rate | 0.01 | 0.045×32/256 |
| Initial Momentum | 0.937 | 0.9 |
| Initial Weight Decay Factor | 0.1 | 0.00004 |
| Batch Size | 16 | 32 |
| Epochs | 300 | 100 |

### 2.2 Two-stage Based Pain Identification

The purpose of this paper is to use the facial features of each sheep to identify whether the sheep is in pain. Fig.1 shows that the recognition method is split into two combined phases: the sheep face detection phase and the sheep face pain detection phase. In the initial step, sheep faces in the flock identified using the lightened yolov5 model, and then the sheep face images are cropped. The second step uses MobileNetV2 to extract feature vectors from cropped sheep facial images for the purpose of pain recognition, in order to avoid the interference of complex backgrounds and sheep breeds on pain recognition results.

## 3 Yolov5 Model Improvement

In this paper, the yolov5 network is lightened in the process of training a dataset of sheep faces using Yolov5. First, the network structure of yolov5s is optimized by combining the structure of Ghostnet to speed up training speed and reduce computation parameters. The structure of BiFPN is then used for feature fusion and ODConv joining to improve detection accuracy. The optimized lightweight Yolov5 network model improved in this study is finally proposed. The specific structure is shown in Fig.2.

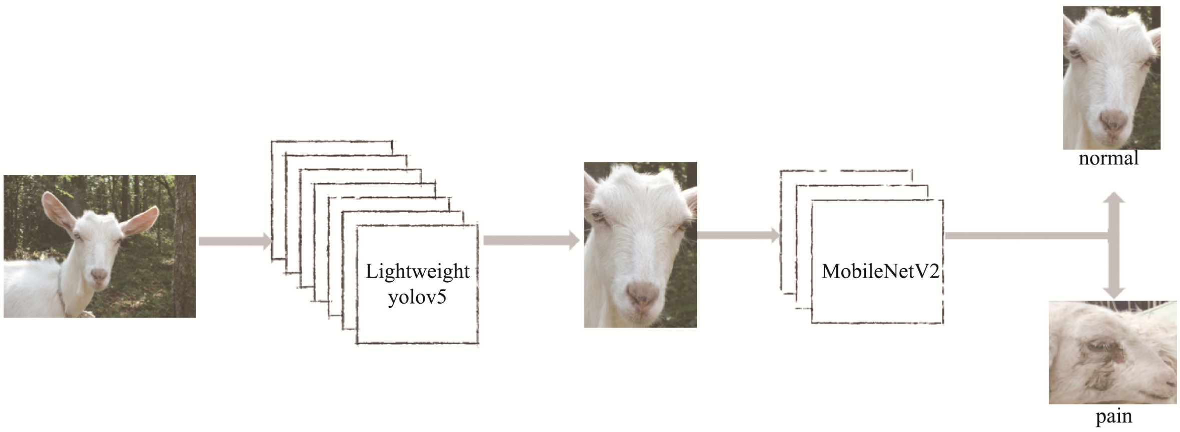



**Fig.1 Sheep Face Pain Detection Method**

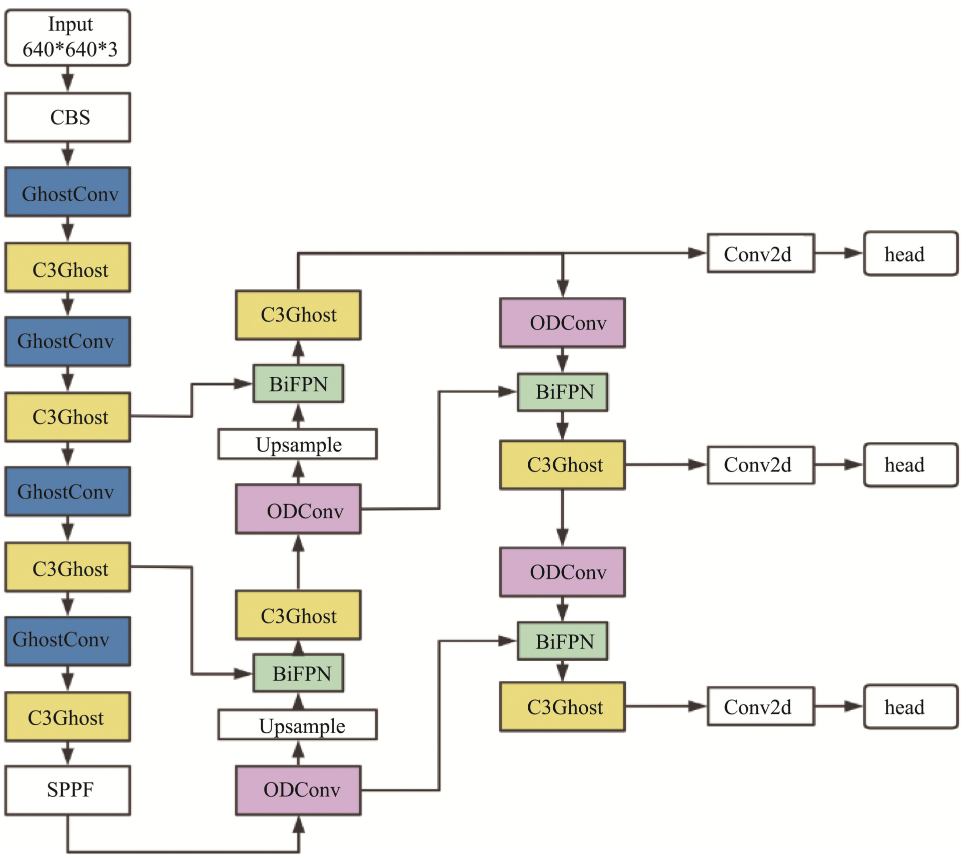



**Fig.2 Improved Yolov5 Network Architecture**

### 3.1 Ghost Module

As the feature maps in the sheep image dataset chosen for this study contain redundant information, the validity of the input data grows as, the size of the model increase. Therefore, this paper combines the ghost module structure in Ghostnet[21] to reduce the computational burden and complexity of the model. First, the base feature map is created via standard convolution. Next, the convolutional linear transformation $\varphi_k$ is performed to obtain the ghost feature diagram and finally, the prior features are superimposed on the output. This method has effectively generated more feature maps with fewer parameters and calculations The structure of the ghost module is shown in Fig.3.

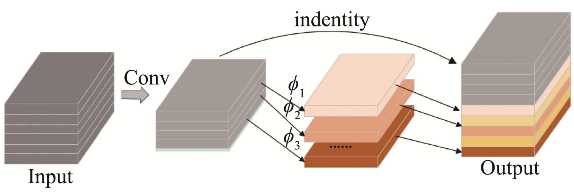



**Fig.3 Ghost Module Structure Diagram**

Furthermore, Fig.4 shows the standard convolution and the Ghost module.

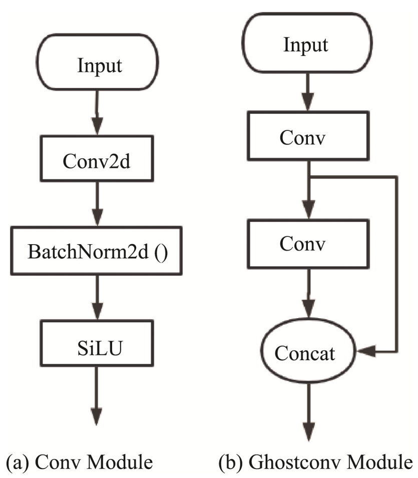



**Fig.4 Conv Module and Ghostconv Module**

Assume that the size of the input feature map is h×$w$×$c$, the size of the output feature map is $h'$×$w'$×$n$, and the size of the convolution kernel is $k$×$k$, where h and $w$ represent the height and width of the input feature map, $h'$ and $w'$ represent the height and width of the output feature map, Respectively, for ordinary convolution, the required number of FLOPs can be calculated as $h'$×$w'$×$n$×$c$×$k$×$k$. The Ghost module consists of a constant mapping $\mathrm{m}\cdot(s-1)=\frac{\mathrm{n}}{s}\cdot(s-1)$ which contains a number of linear operations, with the kernel size of each linear operation being $d$×$d$. The theoretical ratio can be computed by comparing the ghost module to the FLOPs of ordinary convolution as follows:

$$r_s=\frac{n\cdot h'\cdot w'\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot h'\cdot w'\cdot c\cdot k\cdot k+(s-1)\cdot\frac{n}{s}\cdot h'\cdot w'\cdot d\cdot d}=\frac{\mathrm{c}\cdot k\cdot k}{\frac{1}{s}\cdot c\cdot k\cdot k+\frac{s-1}{s}\cdot d\cdot d}\approx\frac{s\cdot c}{s+c-1}\approx s \quad (1)$$

where the magnitude of $d$×$d$ is similar to the magnitude of $k$×$k$ and $s\ll c$. Similarly, the compression ratio of the parametric model can be calculated as:

$$r_c=\frac{n\cdot c\cdot k\cdot k}{\frac{n}{s}\cdot c\cdot k\cdot k+(s-1)\cdot\frac{n}{s}\cdot d\cdot d}\approx\frac{s\cdot c}{s+c-1}\approx s \quad (2)$$

According to formula 2, the number of FLOPs and parameters for standard convolution is around s times that of the ghost module. Hence, both the ghosconv and C3Ghost modules are constructed based on the ghost Module. Reduce the computational cost by substituting ghostconv for the normal convolution and C3Ghost for the original C3 module. C3Ghost is shown in Fig.5.

### 3.2 The Bidirectional Feature Pyramid Network

Yolov5 neck layer achieves feature aggregation by PANet[22] architecture. PANet provides a bottom-up inter-layer transmission path, and the Predicting feature layer enhances target recognition accuracy by combining shallow and deep features and extracting features. However, in the detection process, since different input features have different resolutions, different scale feature layers have different contributions.

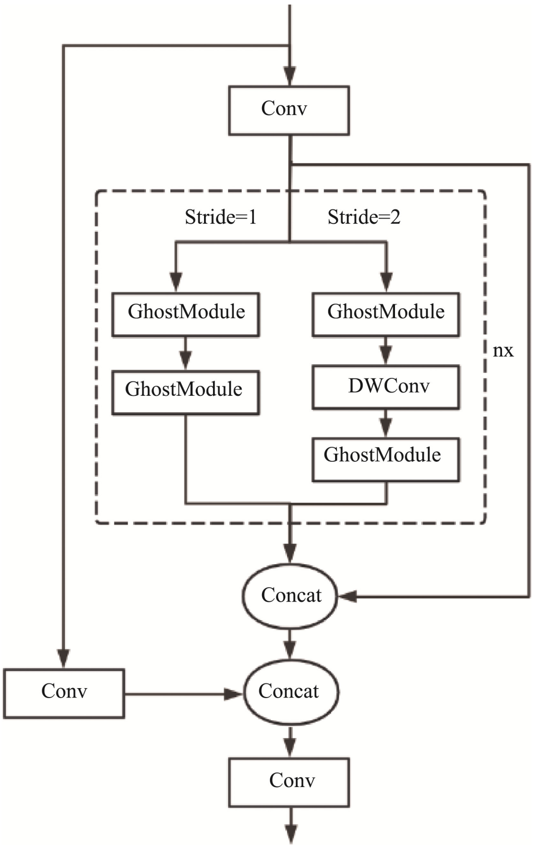



**Fig.5 C3Ghost Structure Diagram**

Therefore, this paper uses the BiFPN[23] structure to replace the original PANet structure, and the fusion process is shown in Fig.6. BiFPN uses context information and weight information to balance different dimensions based on PANet, to achieve simple and fast multi-scale feature fusion by adding edges. Fast normalized fusion is then used to achieve feature layer

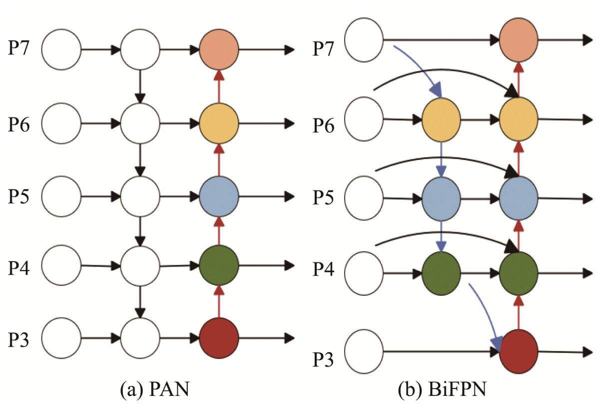



**Fig.6 PAN and BiFPN Structure**

weighting. The fusion process is as follows:

$$o = \sum_{i} \frac{\omega_i * I_i}{\varepsilon + \sum_{j} \omega_j} \tag{3}$$

Where $\omega$ denotes the weight parameter that determines the significance of additional features.

### 3.3 ODConv

To avoid the reduction of the feature extraction capability of the light-weight network, we introduce dynamic full-dimensional convolution to improve the feature extraction and further enhance the network detection accuracy. ODConv[24] is referred to as full-dimensional dynamical convolution because dynamics on the dimensions, i.e., the null field, the input channel, and the output channel are considered simultaneously. Complementary attention learning via a parallel strategy using a multi-dimensional attention mechanism along four dimensions of kernel space. The formula is shown in Equation 4 and ODConv is shown in Fig.7.

$$\begin{aligned} y = (&\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot w_1 + \cdots \\ &+ \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot w_n) * x \end{aligned} \tag{4}$$

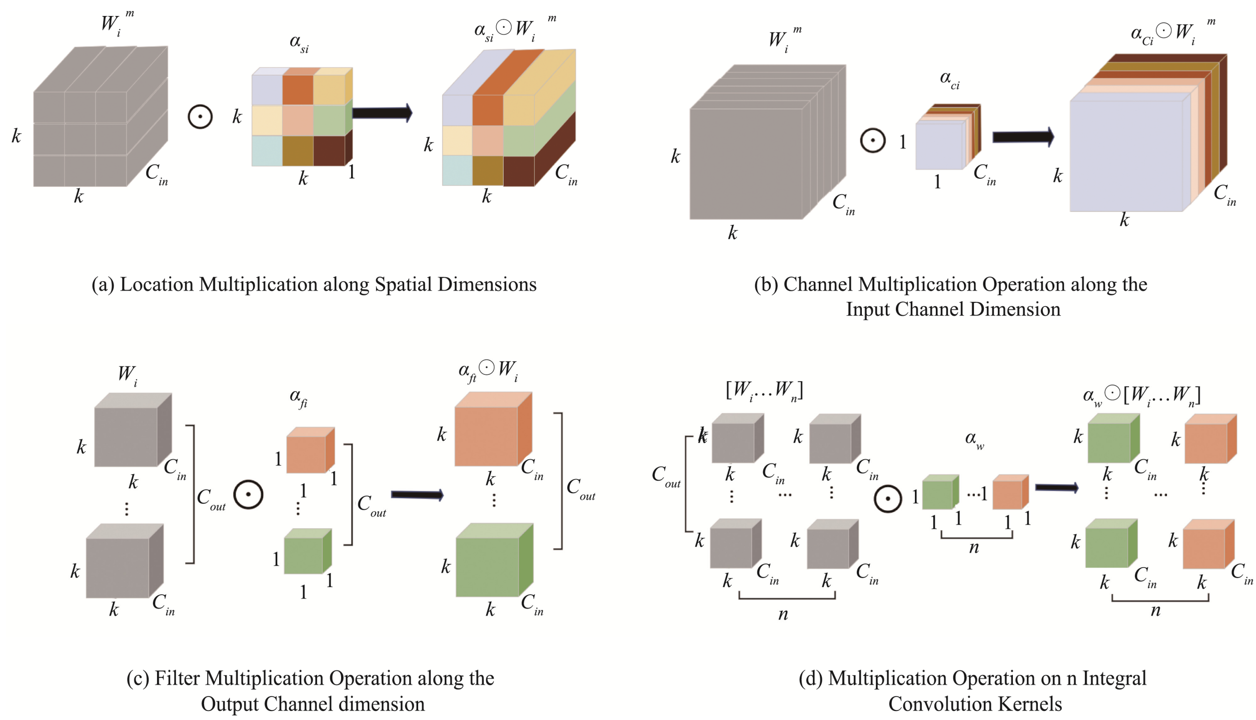
where $\alpha_{w1}$ is the scalar of attention from the convolution kernels, $\alpha_{si} \in R^{k\times k}$, $\alpha_{ci} \in R^{c_{in}}$, $\alpha_{fi} \in R^{c_{out}}$ represent three new attention points introduced on spatial dimensions, input channel dimensions and output channel dimensions, respectively.

The four types of attention described above are complementary and can capture rich contextual information by incrementally multiplying the convolution with different attention along the position, channel, filter and kernel dimensions. As a result, the ordinary convolution in the yolov5 head is replaced with ODConv, and Lightweight network detection accuracy is enhanced by ODConv's powerful convolutional feature extraction capability.

## 4 Experiment and Analysis

### 4.1 Datasets and Pre-processing

Yolov5s dataset is from Billah M[20] and open-source datasets are from Pixabay VEER and Baidu. The total number of images is 1848.



(a) Location Multiplication along Spatial Dimensions

(b) Channel Multiplication Operation along the Input Channel Dimension

(c) Filter Multiplication Operation along the Output Channel dimension

(d) Multiplication Operation on n Integral Convolution Kernels

**Fig.7 Schematic Representation of Four Stepwise Attentional Multiplication of Convolution Kernels in ODConv**

For MobileNetV2 network: This dataset comes from the yolov5 dataset, which is first cleaned by manual inspection to exclude some images of sheep faces with blurred features. Next, data enhancement is performed to increase the number of datasets and avoid the overfitting of the model, as shown in Fig.8.

Sheep's pain was rated according to SPFES recommendations, such as Eyes wide open 0 points, half closed 1 point, all closed 2 points. The shape of the nose is 0 points for "U," 2 points for "V" and 1 point for "U" and "V". The above rules were then used to assess the eyes, noses, cheek and lips in four areas. If the combined score is greater than 1.5, sheep are considered to be in pain[9]. The dataset was divided into normal and outliers based on scores.

(a) Raw Images (b) Changes in Brightness (c) Flip (d) Rotation

**Fig.8 Data Enhancement**

**Table 2 Sheep Face Image Scoring Results**

| Rating/ Points | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Quantity/ Frame | 497 | 44 | 180 | 26 | 10 | 6 | 2 | 0 | 0 |

Note: According to the scoring result, the situation with a score that is less than 1.5 is considered as normal sheep and those with a score greater than 1.5 as pain sheep.

## 4.2 Evaluation Metric

In this paper precision (P), recall (R), F1 value, and average precision (AP) were used as the evaluation metrics for the model's performance, The specific formulas are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{5}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{6}$$

$$F1 = \frac{2 \times P \times R}{P + R} \times 100\% \tag{7}$$

$$AP = \int_0^1 P(R) d(R) \tag{8}$$

where TP (true positive) means that the sample is correctly segregated into positive samples; FP (false positive) indicates that the sample is incorrectly classified as positive samples; FN (false negative) indicates that the sample is incorrectly classified as negative samples, representing the sum of all category numbers.

## 4.3 Ablation Experiments

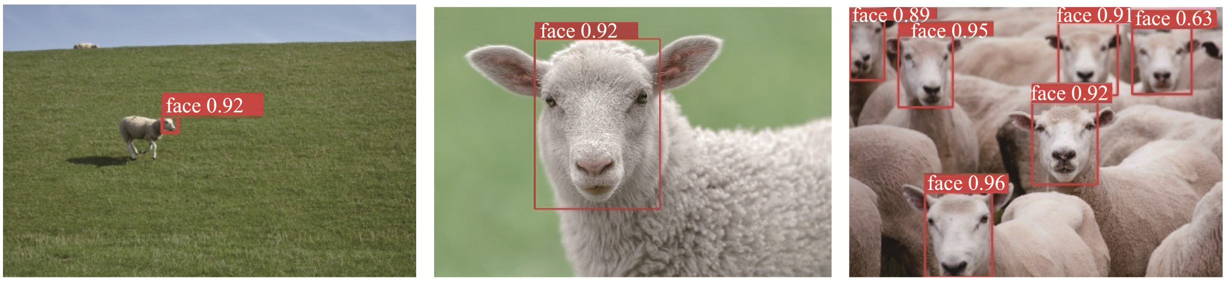
To verify the efficacy of the lightweight yolov5 developed in this paper, ablation experiments were done on the same dataset and the results are shown in Table 3. As shown in Table 3, after first ensuring that the backbone network was modified and the network was made lighter, and after adjusting the loss function, adding BiFPN, and ODConv after various combinations for experimental comparison, the improved model's average accuracy and recall are shown to be optimal. Note that although its accuracy is slightly lower than that of test 4, its size is smaller, and its overall performance is better. The obtained results demonstrate that the enhancements presented in this paper have a positive impact on the model, achieving a reduction in the network's weight without diminishing the average detection accuracy.

The recognition effect of lightweight yolov5s at long range, close range, and in group living is shown in Fig.9. It is discovered that the lightweight yolov5 proposed in this paper has a good detection effect on sheep faces in different scenes, which can address the sheep face detection problem in the sheep breeding process and provide a theoretical foundation for the subsequent sheep face pain detection.

**Table 3 Ablation Experiments**

| Number | Ghostnet | SIoU | BiFPN | ODConv | AP (%) | P (%) | R (%) |
|---|---|---|---|---|---|---|---|
| 1 | √ | × | × | × | 80.1 | 90.4 | 89.0 |
| 2 | √ | √ | × | × | 81.3 | 91.7 | 89.0 |
| 3 | √ | × | √ | × | 81.2 | 90.0 | 90.0 |
| 4 | √ | × | × | √ | 79.9 | 93.9 | 89.0 |
| 5 | √ | √ | √ | × | 81.0 | 87.2 | 90.0 |
| 6 | √ | × | √ | √ | 81.8 | 90.2 | 89.0 |
| Ours | √ | √ | √ | √ | 82.7 | 92.3 | 90.0 |

Note: The bolded data is the optimal value, "√" indicates use of the policy, and "×" indicates non-use of the policy.




(a) Long Range (b) Close Range (c) Flock

**Fig.9 Sheep Face Test Results**

### 4.4 Algorithm Comparison Experiment

To better assess the lightweight yolov5 model for face detection of sheep, a variety of structured network models: yolo7, yolov7 tiny, yolov5s, Faster R-CNN, SSD, and optimized lightweight yolov5 network models were chosen and tested for comparison in this paper, and the results of the comparison are shown in Table 4.

Performance of the light-weight yolov5 model for sheep face detection was evaluated by comparing precision (P), recall (R), F1 score (F1), AP ,and model parametric number and model complexity (GFLOPs) of different models, and the results are shown in Table 3. The analysis reveals that the optimized model has 4.35M parameters and a model complexity of 7.8 when compared to the yolov4 tiny, SSD, Faster R-CNN, and yolov5s networks. The number of parameters was reduced by 25.89%, 83.46%, 84.63%, and 38.03%, while the complexity of the model was reduced by 51.85%, 87.58%, 99.18%, and 50.93%, respectively, indicating that the size of the optimized model was reduced and the recognition accuracy was higher than with other network models. The optimized yolov5 model has 92.3% accuracy, 90.0% recall, 82.8% average accuracy AP, and 81% F1 for sheep face detection, which corresponds to an improvement of 0.2% in precision and 1% in recall over the original model while the average precision remains nearly the same. The above analysis shows that the method proposed in this study can maintain a high detection accuracy while greatly reducing the size of the model, and may therefore address the needs of sheep face detection in complex environments.

## 5 Pain Detection

MobileNetV2 for pain detection of sheep faces recognized by lightweight yolov5 is applied, and Fig.10 illustrates the detection accuracy. The normal sheep face recognition AP value reaches 98% and the painful sheep face recognition AP value is at 99%.

To investigate the detection effect of the model, sheep faces of different ages and breeds were selected from images outside of the training dataset for the purpose of prediction, and the prediction results are shown in Fig.11.

**Table 4 Comparison of Different Networks**

| Model | AP(%) | P(%) | R(%) | F1(%) | Parameters(M) | GFLOPs |
|---|---|---|---|---|---|---|
| Yolov4 Tiny | 72.0 | 88.2 | 56.0 | 70.0 | 5.9 | 16.2 |
| SSD (VGG16) | 78.7 | 94.3 | 64.3 | 76.0 | 26.3 | 62.8 |
| Faster R-CNN (Resnet50) | 86.1 | 61.2 | 86.0 | 72.0 | 28.3 | 948.1 |
| Yolov5s | 82.8 | 92.1 | 89.0 | 81.0 | 7.0 | 15.9 |
| Ours | 82.7 | 92.3 | 90.0 | 80.0 | 4.4 | 7.8 |

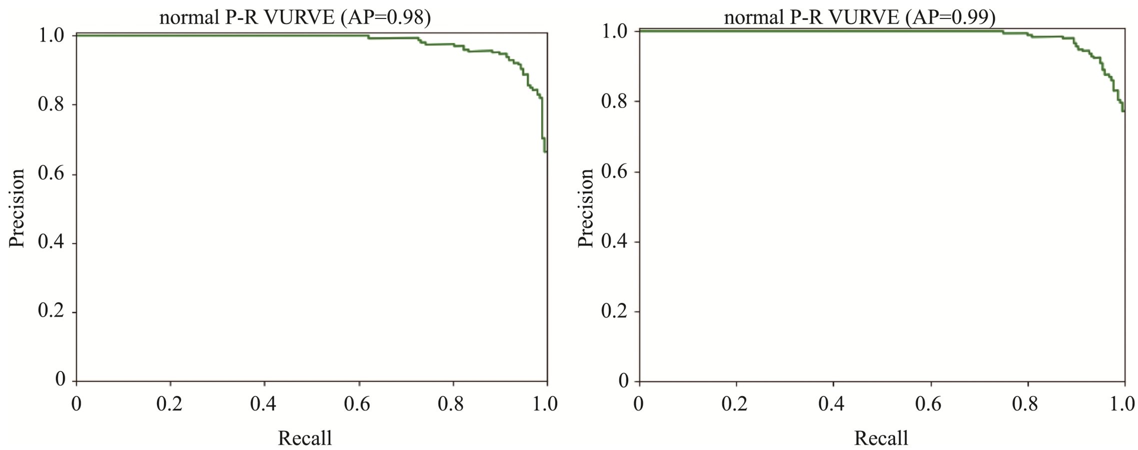



**Fig.10 Model Training Results**


(a) Predicted as Normal Sheep Face Image (b) Predicted as Normal Sheep Face Image

**Fig.11 Examples of Forecast Results**

This indicates that MobileNetV2 has good pain recognition effect on cropped sheep face images by lightweight yolov5. The two images in the figure on pain identification both have half closed eyes, and tight cheeks, which are within the range of pain according to the introduction in section 3.1 on the pain portion of the sheep's face, and the MobileNetV2 network makes accurate predictions on this, while avoiding interference from complex background and breed of sheep etc.

## 6 Conclusion

A lightweight model with yolov5s is proposed as the base model and the original base network combined with Ghostnet in order to reduce the weight of the network. Using the SIoU loss function, the bidirectional feature pyramid network and ODConv are introduced to improve the detection accuracy of the lightweight network and to accelerate the convergence of the network.

A two-stage recognition algorithm is proposed which allows pain recognition in complex contexts and for a variety of sheep breeds. For sheep face detection, the lightened network has a precision of 92.3%, recall of 90.0%, AP of 82.7%, and F1 of 80.0%, corresponding to a 38.03% decrease in the number of parameters and a 50.94% decrease in model complexity compared to yolov5s, whereas the precision is improved by 0.2% and the recall is improved by 1%. As the model size is reduced, the accuracy is taken into account, allowing for the recognition and cropping of sheep faces while avoiding the impacts of sheep breeds and complex backgrounds in the recognition results. MobileNetV2 achieved an average accuracy of 98% and 99%, respectively, for normal and painful sheep, and achieved good results for pain recognition of sheep faces after the image cropping of lightweight yolov5.

The results demonstrate that the proposed two-stage sheep face pain recognition method based on lightweight yolov5 can be used in sheep health farming.

## References

[1] McLennan K M. Why pain is still a welfare issue for farm animals, and how facial expression could be the answer[J]. Agriculture, 2018, 8(8): 127.

[2] Langford D J, Bailey A L, Chanda M L, et al. Coding of facial expressions of pain in the laboratory mouse[J]. Nature methods, 2010, 7(6): 447-449.

[3] Keating S C J, Thomas A A, Flecknell P A, et al. Evaluation of EMLA cream for preventing pain during tattooing of rabbits: changes in physiological, behavioural and facial expression responses[J]. 2012.

[4] Dalla Costa E, Minero M, Lebelt D, et al. Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration[J]. PLoS one, 2014, 9(3): e92281.

[5] Flecknell P. Analgesia from a veterinary perspective[J]. British Journal of Anaesthesia, 2008, 101(1): 121-124.

[6] Dolan S, Kelly J G, Monteiro A M, et al. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia[J]. Pain, 2003, 106(3): 501-512.

[7] Dolan S, Field L C, Nolan A M. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation[J]. Pain, 2000, 86(3): 311-320.

[8] Marteniuk J V, Herdt T H. Pregnancy toxemia and ketosis of ewes and does[J]. Veterinary Clinics of North America: Food Animal Practice, 1988, 4(2): 307-315.

[9] McLennan K M, Rebelo C J B, Corke M J, et al. Development of a facial expression scale using footrot and mastitis as models of pain in sheep[J]. Applied Animal Behaviour Science, 2016, 176: 19-26.

[10] Wang Z, Liu T. Two-stage method based on triplet margin loss for pig face recognition[J]. Computers and Electronics in Agriculture, 2022, 194: 106737.

[11] Kleanthous N, Hussain A J, Khan W, et al. A survey of machine learning approaches in animal behaviour[J]. Neurocomputing, 2022, 491: 442-463.

[12] Bao J, Xie Q. Artificial intelligence in animal farming: A systematic literature review[J]. Journal of Cleaner Production, 2022, 331: 129956.

[13] Meng X, Tao P, Han L, et al. Sheep Identification with Distance Balance in Two Stages Deep Learning[C]//2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC). IEEE, 2022, 6: 1308-1313.

[14] Hummel H I, Pessanha F, Salah A A, et al. Automatic pain detection on horse and donkey faces[C]//2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). IEEE, 2020: 793-800.

[15] Lu Y, Mahmoud M, Robinson P. Estimating sheep pain level using facial action unit detection[C]//2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE, 2017: 394-399.

[16] Noor A, Zhao Y, Koubâa A, et al. Automated sheep facial expression classification using deep transfer learning[J]. Computers and Electronics in Agriculture, 2020, 175: 105528.

[17] Yu L, Pu Y, Cen H, et al. A Lightweight Neural Network-Based Method for Detecting Estrus Behavior in Ewes[J]. Agriculture, 2022, 12(8): 1207.

[18] Song S, Liu T, Wang H, et al. Using Pruning-Based YOLOv3 Deep Learning Algorithm for Accurate Detection of Sheep Face[J]. Animals, 2022, 12(11): 1465.

[19] Li X, Du J, Yang J, et al. When Mobilenetv2 Meets Transformer: A Balanced Sheep Face Recognition Model[J]. Agriculture, 2022, 12(8): 1126.

[20] Billah M, Wang X, Yu J, et al. Real-time goat face recognition using convolutional neural network[J]. Computers and Electronics in Agriculture, 2022, 194: 106730.

[21] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.

[22] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[23] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[24] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.

## Author Biographies

**LI Huan** is now a M.Sc. candidate in College of Electrical and Mechanical Engineering, Xi'an University of Architecture and Technology. His main research Interests include deep learning and target detection algorithms.

E-mail: 415410533@qq.com

**HE Lile** is now a Professor and the Doctoral Supervisor in College of Electrical and Mechanical Engineering, Xi'an University of Architecture and Technology. His main research interests include robot intelligence technology and machine learning.

E-mail: hllnh2013@163.com