돼지 공격 행동 모니터링을 위한 영상 기반의 경량화 시스템

하싼*, 이종욱**, 오스만*, 박대희**, 정용화**
*고려대학교 컴퓨터정보학과, **고려대학교 컴퓨터융합소프트웨어학과
e-mail: 2017011010@korea.ac.kr

Lightweight Video-based Approach for Monitoring Pigs' Aggressive Behavior

Hassan Seif Mluba*, Jonguk Lee**, Othmane Atif*, Daihee Park**, Yongwha Chung**
*Dept. of Computer Information Science, Korea University
**Dept. of Computer Convergence Software, Korea University

Abstract

Pigs' aggressive behavior represents one of the common issues that occur inside pigpens and which harm pigs' health and welfare, resulting in a financial burden to farmers. Continuously monitoring several pigs for 24 hours to identify those behaviors manually is a very difficult task for pig caretakers. In this study, we propose a lightweight video-based approach for monitoring pigs' aggressive behavior that can be implemented even in small-scale farms. The proposed system receives sequences of frames extracted from an RGB video stream containing pigs and uses MnasNet with a DM value of 0.5 to extract image features from pigs' ROI identified by predefined annotations. These extracted features are then forwarded to a lightweight LSTM to learn temporal features and perform behavior recognition. The experimental results show that our proposed model achieved 0.92 in recall and F1-score with an execution time of 118.16 ms/sequence.

1. INTRODUCTION

Pigs are the source of one of the most consumed meats globally, with their production accounting for over 35% in the livestock industry [1]. This popularity has attracted research work aiming at improving farm productivity and reducing running costs by addressing common issues that affect pig's health and welfare. One of those issues is the occurrence of pigs' aggressive behaviors, which result in injuries and social stress between pigs. These behaviors negatively affect pigs' growth, increase both mortality rates and veterinary cost [2], and eventually lead to a financial burden for pig farmers. For example, tail-biting alone was estimated to cost an average loss of ￡18.96 (≈ ₩ 24,450) per victim, with the number increasing when accompanied with other aggressive behaviors [3]. Despite the efforts made by pig caretakers, continuously monitoring multiple pigs for 24 hours to manually identify those behaviors is a very difficult task. Therefore, there is a need for a practical solution to monitor pigs' aggressive behavior so as to complement the work done by pig caretakers.

Studies using information technology to automatically detect pigs' abnormal behaviors that result in serious problems which affect farms productivity have already been reported. By considering the representative studies, we find that Lee et al. [4] confirmed that by extracting features from images acquired through a Kinect depth sensor and classifying them using Support Vector machines (SVM), it was possible to effectively detect pigs' aggressive behaviors. In addition, Chen et al. [5] recently published a study utilizing deep learning algorithms, as they have proven to significantly improve the performance of systems in various fields, in order to recognize pigs' aggressive episodes. Their proposed system uses a VGG16 model to extract image features from images and then a Long Short-Term Memory (LSTM) model, which can learn long-term dependencies allowing it to perform well on sequential data, to extract temporal features and identify pig behaviors. This approach that applies two deep learning models proved to be effective since it achieved good performance results in pigs' aggressive behavior recognition. However, the use of such method necessitates the installation of a costly high computing environment for its execution, making its implementation in small-scale farms with limited budgets financially unfeasible.

Recently, to address this issue, different lightweight deep learning models have been presented. Among those models is the Mobile neural architecture search Network (MnasNet) developed by Tan et al. [6], which is discovered through the Neural Architecture Search (NAS) technique to guarantee that it is optimized for mobile environments. MnasNet has shown a good classification performance and allows to control the model size and latency further by removing filters in each layer through the Depth Multiplier (DM) parameter. The aforementioned MnasNet has been used by Hong et al. [7] as a lightweight deep learning method to automatically extract sound features from a spectrogram expressing pig sound information as an image. It was also confirmed that the model proposed in [7] can be executed in the embedded board NVIDIA TX-2 making it applicable even in small-scale farms. Additionally, a lightweight LSTM model has been introduced [8-9] to reduce the heavy deep learning structure of LSTM, and this was obtained by examining different parameters to identify the most optimal configuration with the least number of layers and unit that maintain the LSTM's detection performance. It has also been reported that such lightweight LSTM structure allows to use a reduced size of the model while maintaining the good detection performance of LSTM [8]. Thus, in this study, by introducing lightweight deep learning techniques to perform image and temporal features extraction, we propose a pig's aggressive behavior detection system with a reduced model size and faster processing speed that still maintains a good performance.

For the purpose of implementing our system, we first use MnasNet, which was used to automatically and effectively extract image features not in PC environment, but rather in lightweight systems with limited computing power. Then, to extract temporal information and detect the pigs' behavior, we use a lightweight LSTM. The process to obtain the lightweight LSTM was done in accordance with the method applied by Kayode and Tosun [9] to implement a lightweight LSTM on edge devices. The system proposed in this study makes use of our proposed method, which is referred to through this paper as the MnasNet-LightweightLSTM (MN-LLSTM) model, and has been experimentally verified using videos collected from real pigpens.

2. PROPOSED METHOD

The architecture of our proposed method is shown in Figure 1. The system is composed of three main modules: data acquisition module, region of interest (ROI) identification module, and behavior recognition module.

2.1 Data Acquisition Module

The data acquisition module is the part of the system responsible for gathering data. An RGB camera is used to collect video data of several pigs located inside a pigpen. Then, frames are extracted from the video and forwarded to the next module.

2.2 ROI Identification Module

In this module, upon receiving the extracted frames, each three consecutive frames are grouped as a sequence. Then, using the pigs bounding box coordinates and IDs that were manually annotated beforehand, the module identifies the region of interest (ROI) of every individual pig in the sequence. For each frame in the sequence, we retrieve the ROI of all the pigs present and assign to each one its corresponding annotation ID. Consequently, for each sequence, every individual pig will have an ID and three extracted ROIs, one from each of the three consecutive frames. This results in a total of $n \times 3$ ROIs for n pigs present in the frames of a single sequence. The resulting sequences of ROIs are then forwarded to the next module in order of their assigned IDs to perform behavior recognition for each individual pig.

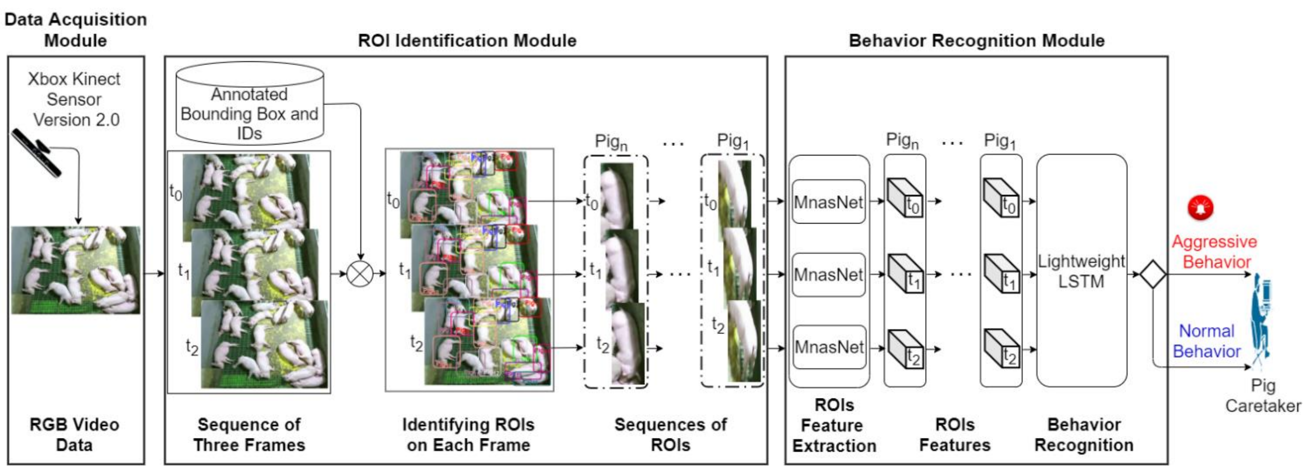
2.3 Behavior Recognition Module

In this module we use the MN-LLSTM model for pig behavior recognition. First, the sequences of identified ROIs received from the previous module are fed sequentially to the MnasNet model to perform image feature extraction. MnasNet is a CNN-based lightweight model defined through NAS for mobile and that uses Depth Multiplier (DM) to further decrease the model size by reducing the number of filters in the network based on the value of the DM. The classifier layer was removed from the MnasNet model in order to use it as a feature extractor. At this point, the system uses individual identified ROIs from each sequence as inputs to MnasNet and rearranges the three extracted feature vectors into a sequence. Each sequence of extracted features is then forwarded to the lightweight LSTM as input where temporal features are extracted to finally recognize the pig's behavior. The LSTM model was used to learn temporal features because of its ability to retain long-term dependencies in sequence data, which makes it a suitable choice for tasks such behavior recognition. Moreover, to obtain a lightweight LSTM network, we investigated the smallest possible number of layers and units that maintained a good recognition performance of the overall MN-LLSTM model.

3. IMPLEMENTATION DETAILS AND EXPERIMENTAL RESULTS

3.1 Data Collection

The video data used in our study was collected using an Xbox Kinect Sensor Version 2.0 (EN/XC/FR/ES AOC HD) mounted about 3m from floor in a pig farm located in Sejong City. The pigpen where the videos were recorded contained a total of 17 pigs. The data was collected for 24 hours in some days to capture more samples of low-frequency behaviors displayed by pigs such as ear in the mouth, being ear bitten,



(Figure 1) Proposed architecture for lightweight pig aggressive behavior monitoring system.

tail in the mouth, being tail bitten, etc. Thereafter, video clips containing the targeted behaviors were selected and annotated using Video Tracking and Behavior Annotation Tool (ViTBAT) [10] to prepare both bounding box annotations and behavior annotations of each individual pig. The generated bounding boxes with their IDs are the ones used to identify individual pigs' ROIs in the ROI identification module.

A total of 89 video clips were annotated, with each video having an average duration of 10 seconds at a rate of 30 frames per second. The frames were extracted from each video clip, then grouped into sequences of three successive frames. Each frame has a size of 960 in width and 540 in height. A total of 13 observed behaviors with a varying number of occurrences were used to create the dataset.

The dataset used in the experiments was derived from multiple videos containing different pigs labeled each with an ID, a bounding box, and one of 13 behaviors. In order to maintain and map each ROI ID with their corresponding behavior sequences across the consecutive frames, the dataset was created by splitting video clips into clips used for training and clips used for testing. Accordingly, we selected 68 video clips for the training set and 21 video clips for the testing set. The selection of videos was carefully done to ensure a balanced distribution of each behavior across the training and testing set. The entire dataset contains a total of 144466 sequences. From those sequences, 110466, which accounts for 76.5% of the dataset, were used as a training set while the remaining 23.5%, that is 34000 sequences, were used as a testing set. Table 1 shows the dataset used in our experiment in terms of number of sequences of three frames for each behavior.

<Table 1> Pig behavior dataset.

Category	Behavior	Number of sequences
Normal	Eating or drinking	20394
	Sleeping	43148
	Sitting or resting	29542
	Moving	15390
Aggressive	Nose in the belly	2871
	Being belly nosed	2239
	Tail in the mouth	2787
	Being tail bitten	2749
	Ear in the mouth	2217
	Being ear bitten	2055
	Head knocking the body	8101
	Body being knocked by head	7186
	Head being knocked	5787

3.2 Implementation Details

All the experiments were carried out in a desktop computer running Windows 10 with a 3.0 GHz Intel Core i5 CPU, 32 GB RAM, and an Nvidia GeForce GTX 1080Ti graphics card. The implementations and trainings of the models were performed using Python 3.6 and Pytorch 1.1.

Experiments using the MN-LLSTM model and a VGG16-LSTM model were conducted for performance comparison. Although the VGG16-LSTM was based on the method used by Chen et al. [5], the network configurations and parameters were defined differently to fit our dataset requirements. The input size of MnasNet is 224×224 and it outputs features of size 1280, whereas VGG16 has an input size of 224×224 and outputs features of size 4096. Following the analysis of results obtained from different parameters of the lightweight LSTM, we concluded that a single layer with 64 hidden units is the smallest size of LSTM that maintains its performance. Hence, the lightweight LSTM model was implemented with a single-layer of 64 hidden units and an input size of $1 \times 3 \times 1280$ in the MN-LLSTM, and with an input size of $1 \times 3 \times 4096$ in the VGG16-LSTM. A total of three experiments were conducted including one using the VGG16-LSTM model and two using MN-LLSTM with different DM values in MnasNet, namely 1.0 and 0.5 respectively. The DM value represents the rate at which the filters were maintained in MnasNet before training the MN-LLSTM model. DM 1.0 represents the base MnasNet without any filter pruning and DM set to 0.5, as used in our proposed method, is used to prune the MnasNet filters at a rate of 50%.

As for the training, cross-entropy was used as a loss function and Stochastic Gradient Descent (SGD) as an optimizer for all models with the gamma value set as 0.1 and the epsilon as 1exp-9. Our proposed model was trained with a learning rate of 0.005 and the other models were trained with a learning rate of 0.001. The batch size was set to 51, and the number of training epochs was 50 in all the experiments.

3.3 Experimental Results and Comparison

The experimental results for pig behavior recognition are presented in Table 2. The table represents the results of the experiment using VGG16-LSTM and two experiments with MN-LLSTM using different DM values in MnasNet. The experimental results show that our proposed method achieved a recall and F1-score of 0.92. In addition, as detailed in Table 3, setting the DM value to 0.5 helped reduce the number of parameters to 2,922,613, and the execution time to 118.16 ms/sequence. This proves that our proposed method has 98% fewer parameters and an execution time five times faster than VGG16-LSTM. Besides, despite considerably reducing the number of parameters and size of the model, it has maintained a good performance with a drop of only 0.01 in recall and F1-score compared to VGG16-LSTM.

4. CONCLUSION

The occurrence of pig aggressive behavior in pigpens negatively affects pigs' health and welfare resulting in a financial burden to pig farmers. Manually monitoring several pigs for 24 hours to identify those behaviors is a very difficult task for pig caretakers. Thus, to provide a solution that is affordable and applicable even in small-scale farms, in this study, we propose a lightweight model for monitoring pigs' aggressive behavior. The proposed method uses MnasNet (DM 0.5) as an image feature extractor with a lightweight LSTM for pigs' behavior recognition. The results showed that our proposed method has maintained its behavior detection performance despite the model having 98% parameters less than the VGG16-LSTM model, which assures a higher possibility of successful implementation even in low-computing environments.

In our future work, we intend to introduce an automatic detection and tracking identification module in our system.

<Table 2> Comparison of pig behavior recognition experimental results.

Pig Behavior	VGG16-LSTM		MN (DM 1.0)-LLSTM		MN (DM 0.5)-LLSTM		Support
	Recall	F1-score	Recall	F1-score	Recall	F1-score	
Eating or drinking	**0.99**	**0.96**	0.97	**0.96**	0.97	**0.96**	4402
Sleeping	**0.99**	**0.98**	0.98	**0.98**	0.98	**0.98**	11525
Sitting or resting	**0.94**	**0.93**	0.90	0.92	0.91	**0.93**	6548
Moving	0.86	**0.86**	**0.88**	0.83	0.87	0.84	3058
Nose in the belly	0.93	0.94	0.93	0.96	**0.94**	**0.97**	939
Being belly nosed	0.97	0.96	0.97	0.97	**1.00**	**0.98**	639
Tail in the mouth	0.94	0.93	**0.96**	0.92	0.95	**0.94**	1061
Being tail bitten	0.91	**0.94**	0.92	0.92	0.89	0.89	1058
Ear in the mouth	0.74	0.76	0.77	**0.82**	**0.82**	0.81	455
Being ear bitten	0.76	0.83	0.83	0.87	**0.84**	0.87	460
Head knocking the body	**0.71**	**0.76**	**0.71**	0.73	**0.71**	0.74	1618
Body being knocked by head	0.72	**0.75**	**0.75**	0.74	0.74	0.72	1299
Head being knocked	0.80	0.78	**0.82**	**0.79**	0.79	0.75	938
Weighted average	**0.93**	**0.93**	0.92	0.92	0.92	0.92	-

<Table 3> Comparison of number of parameters and execution time in pigs' behavior recognition models.

	VGG16-LSTM	MN (DM 1.0)-LLSTM	MN (DM 0.5)-LLSTM
Number of parameters	135,335,309	5,087,413	**2,922,613**
Execution time (ms/sequence)	645.88	228.07	**118.16**

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1I1A3070835 and NRF-2021R1I1A3049475) and by BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).

References

[1] H. Ritchie and M. Roser, "Meat and Dairy Production - Our World in Data," 2019. Available online: https://ourworldindata.org/meat-production (accessed on 08 November 2020).

[2] J. Marchant-Forde and R.M. Marchant-Forde, "Minimizing Inter-pig Aggression During Mixing," Pig News and Information, Vol. 26, No. 3, pp. 63-71, 2005.

[3] R.B. D'Eath, J.K. Niemi, B.V. Ahmadi, K.M.D. Rutherford, S.H. Ison, S.P. Turner, H.T. Anker, T. Jensen, M.E. Busch, K.K. Jensen, A.B. Lawrence, and P. Sandoe, "Why are Most EU Pigs Tail Docked? Economic and Ethical Analysis of Four Pig Housing and Management Scenarios in the Light of EU Legislation and Animal Welfare Outcomes", Animal, Vol. 10, No. 4, pp. 687-699, 2016.

[4] J. Lee, L. Jin, D. Park, and Y. Chung, "Automatic Recognition of Aggressive Behavior in Pigs Using a Kinect Depth Sensor," Sensors (Switzerland), Vol. 16, No. 5, 631, 2016.

[5] C. Chen, W. Zhu, J. Steibel, J. Siegford, K. Wultz, J. Han, and T. Norton, "Recognition of Aggressive Episodes of Pigs Based on Convolutional Neural Network and Long Short-term Memory," Computer and Electronics in Agriculture, Vol. 169, 105166, 2020.

[6] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and V. Le, "Mnasnet: Platform-aware Neural Architecture Search for Mobile," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, California, USA, pp. 2820-2828, 2019.

[7] M. Hong, H. Ahn, O. Atif, J. Lee, D. Park, and Y. Chung, "Field-applicable Pig Anomaly Detection System Using Vocalization for Embedded Board Implementations," Applied Sciences, Vol. 10, No. 19, 6991, 2020.

[8] M. Jeong, M. Park, J. Nam, and B.C. Ko, "Light-weight student LSTM for real-time wildfire smoke detection," Sensors, Vol. 20, No. 19, 631, 2020.

[9] O. Kayode and A.S. Tosun, "LIRUL: A Lightweight LSTM Based Model for Remaining Useful Life Estimation at the Edge," Proceedings of the IEEE International Computer Software and Applications Conference (COMPSAC), Milwaukee, WI, USA, pp. 177-182, 2019.

[10] T.A. Biresaw, T. Nawaz, J. Ferryman, and A.I. Dell, "ViTBAT: Video Tracking and Behavior Annotation Tool," IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Colorado Spring, USA, pp. 295-301, 2016.