GAN 기반의 영상 잡음에 강인한 돼지 탐지 시스템

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

GAN-based Video Denoising for Robust Pig Detection System

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

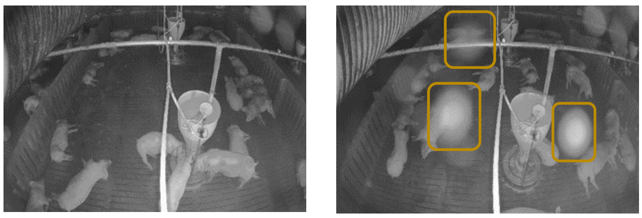
Abstract

Infrared cameras are widely used in recent research for automatic monitoring the abnormal behaviors of the pig. However, when deployed in real pig farms, infrared cameras always get polluted due to the harsh environment of pig farms which negatively affects the performance of pig monitoring. In this paper, we propose a real-time noise-robust infrared camera-based pig automatic monitoring system to improve the robustness of pigs' automatic monitoring in real pig farms. The proposed system first uses a preprocessor with a U-Net architecture that was trained as a GAN generator to transform the noisy images into clean images, then uses a YOLOv5-based detector to detect pigs. The experimental results show that with adding the preprocessing step, the average pig detection precision improved greatly from 0.639 to 0.759.

1. INTRODUCTION

Pork is the most consumed meat in Korea and represents the third largest consumption per capita in the world [1]. It is primarily produced in intensive pig farms in Korea, and with the continuous expansion of farms, the number of pigs is increasing annually [2]. The large number of pigs compared to the limited number of staff at farms makes it hard to maintain the health and welfare of pigs, which affects the efficiency of pig farm. Compared to the traditional pig farming method of managing pigs' health and welfare condition manually by the staff, pig's automatic monitoring provides an efficient way to monitor pigs' health condition without the need for many staff members. This can also help with diagnosing pigs' diseases at an early stage, which would decrease pigs' mortality rates and eventually increase the pork production ability to satisfy the growing demand for pork.

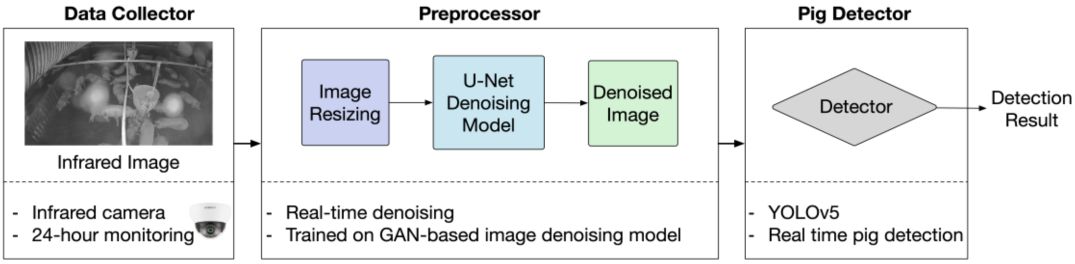
Among the pigs' automatic monitoring methods, infrared camera-based methods not only provide a non-invasive and intuitive way for pig monitoring, but can also work effectively in low light conditions. Since most farms turn off their lights at night, such methods have been getting more attention in recent research work [3-4]. However, when installed in a real farm, infrared cameras always get polluted by the harsh environment of the pig farm, and this negatively affects the pig monitoring systems. For example, with the large number of flying insects present in the farms, there are always cases when insects produce secretion on the camera lens. Those contaminants formed on the camera cause the IR reflection problem and form haze spots on the captured frames, which then significantly affects the pig monitoring performance. Figure 1 shows an example of a clean image and a noisy image both captured by the same camera. New haze spots will form on a camera every few days and with the large number of cameras that are used to monitor all the pig pens in a farm, it is unpractical to regularly clean every camera manually. To improve the robustness of the pig monitoring system and maintain the pig detection performance, it is necessary to remove those haze spots that appear in the noisy images and get their corresponding clean images.



(a) Clean image (b) Noisy image

(Figure 1) A clean image captured after the camera installation and a noisy image captured by the same camera a few days later.

As seen in Figure 1, the haze spot in the bounding box of the noisy image has properties that are very similar to the haze that appear in the image dehazing problem where a lot of breakthrough work has been done recently. The earlier work on image dehazing mainly relied on hand-crafted features which based on textural, chromatic and contrast properties [5]. Those methods could remove the haze to a certain extent, but their performance depended on the accuracy of the hand-crafted features and they failed at scene generalization. Later, with deep learning algorithms rising in popularity, image dehazing methods started utilizing them to improve the performance results. For instance, Cai et al. [6] used convolutional neural networks (CNNs) to estimate haze relevant features and used them to recover haze-free images.



(Figure 2) Overall pipeline of the proposed system

Although deep learning-based methods outperformed prior-based methods, the two steps of haze feature estimation and haze-free image recovery still result in errors of accumulation which affects the dehazing performance. Recently, motivated by the success of generative adversarial networks (GANs), which consist of two competing neural networks named generator and discriminator to perform image-to-image translation [7], Li et al. [8] proposed to transform the haze image to haze-free image pixel to pixel directly by applying a GAN model to image dehazing. Using a U-Net model, which can build multi-scale feature mapping between the haze image and haze-free image, as the generator, they generated very realistic haze-free images from haze images and outperformed previous studies.

Inspired by recent work about image dehazing, this paper applies a GAN-based method with a U-Net model as a generator for solving the haze spot problem in videos recoded in real pig farms and propose a real-time noise-robust infrared camera-based pig automatic monitoring system. The proposed system first uses a U-Net architecture trained as a generator and a regular CNN architecture trained as a discriminator in a GAN-based method to remove the haze spots in noisy images. Then, a YOLOv5-based [9] object detector, which is one of the best detectors in terms of detection accuracy and inference speed, is used to detect pigs inside a pigpen.

2. PROPOSED METHOD

The overall pipeline of our proposed system is shown in Figure 2. The proposed system mainly consists of three modules: a data collector module, a preprocessor module, and a pig detector module.

2.1 Data Collector

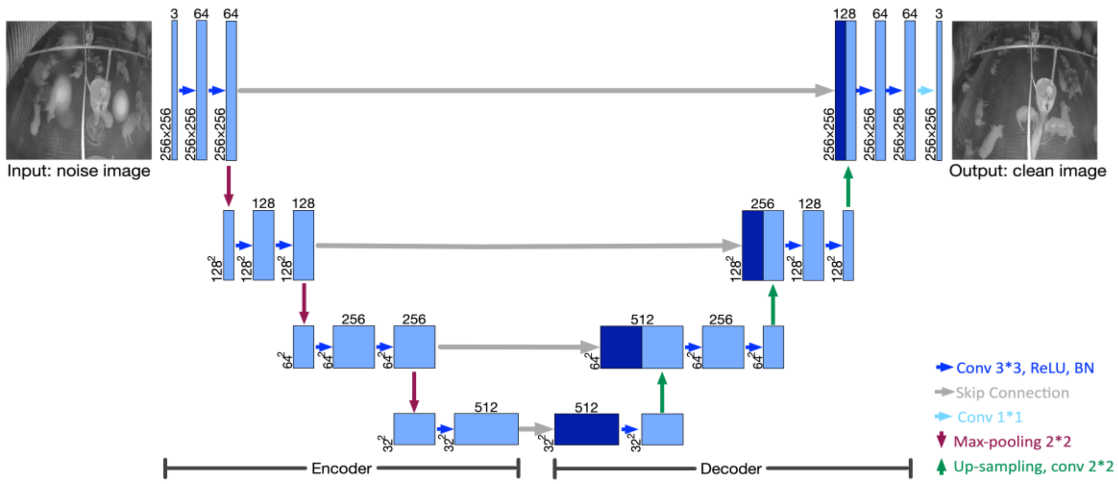
This module is responsible for collecting the infrared images through an infrared camera that can monitor pigs continuously for 24 hours. The camera is installed at the top of a pigpen at an inclined angle.

2.2 Preprocessor

The preprocessor module is where noisy images are transformed into clean images. In this module, noisy images are first resized into 256×256 images, then denoised to obtain clean images using an image denoising model. In order to obtain realistic denoised images, the image denoising model is trained using a GAN-based approach that consists of a generator and a discriminator which are introduced in detail in the following sections.

2.2.1 Generator

The generator, as shown in Figure 3, has a symmetric structure named U-Net which consists of a contractive encoder path, an expending decoder path, and a skip connection between symmetric layers of the encoder and decoder. The input of U-net is a $256 \times 256 \times 3$ noisy image followed by two groups of convolutional layers, ReLU activation, and Batch Normalization (BN) layers to get a $256 \times 256 \times 64$ feature map. Then the feature map is down-sampled and used to extract more abstract features in a deeper layer. Every time the network goes deeper in the encoder path, the size of the feature map is halved whereas the number of feature channels is doubled. A $32 \times 32 \times 512$ feature map is acquired at the end of the encoder path. The operation in the decoder path is an inverse operation of the encoder path, with several convolution and up-sampling operations, it gradually increases the size of the feature map while decreasing the number of feature channels. A $256 \times 256 \times 64$ feature map

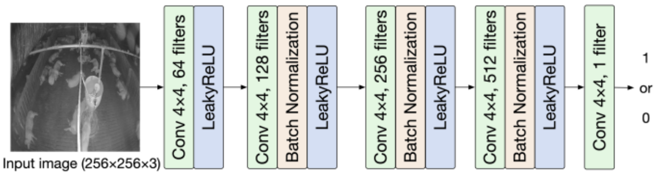


(Figure 3) The generator structure of the image denoising model.

is acquired at the end of the decoder's path. The skip connections between the encoder and decoder copy the features from the encoder and concatenate them with the up-sampled features in the decoder to reduce the features loss when down-sampling the features. Finally, a 1×1 convolutional layer is used to reduce the number of feature maps and recover the target clean image.

2.2.2 Discriminator

The discriminator can be considered as a classifier, and it is used to distinguish whether an input image is a clean image or a denoised image produced by the generator. The structure of the discriminator used, as shown in Figure 4, is a regular CNN where the basic operations are convolution, LeakyReLU activation and batch normalization. During each training iteration clean images and images denoised by the generator are fed to the discriminator, alternatively. The discriminator outputs a number ranging from 0 to 1. When the input is a clean image, the discriminator is trained to output a number as close to 1 as possible, as 1 means that the input is a clean image. On the other hand, when the input image is the denoised image, the discriminator is trained to output a number as close to 0 as possible since 0 means that the input is a denoised image.



(Figure 4) The structure of the CNN model used as discriminator.

2.3 Pig Detector

This module is where the detection of the pigs present in the denoised images is performed. The object detector selected to perform the pigs' detection is YOLOv5 [9] as it is one of the models that achieves state-of-the-art object detection performance and guarantees the best detection accuracy and inference speed.

3. EXPERIMENTAL RESULTS

3.1 Data Collection and Datasets

The data was collected from January 2021 to February 2021 in a pig farm at Hadong-gun, Gyeongsangnam-do, South Korea. The clean images and noisy images were collected by the same infrared camera (QND-6012R, Hanwha Techwin Co.) with a time interval of four weeks. The resolution of the collected images is 1920×1080 with 10 fps.

Since the image denoising model needs to be trained to learn a mapping between a noisy image and a clean image, paired noisy images and clean images are required for the model training. However, due to the constant movement of pigs inside the pigpen, their positions in the collected noisy images and clean images are always different. Thus, it is impossible to pair clean images with real noisy images. To solve this issue, the noisy images needed to train the denoising model were simulated by adding noise to clean images and then pairing them. We simulated the noise with the help of the atmospheric scattering model (Eq. 1) which shows the formation step of the haze image and has been widely used in the early image dehazing task [5-6].

$$I(x) = J(x)t(x) + A\big(1 - t(x)\big) \tag{1}$$

Here, $J(x)$ is the clean image, $I(x)$ is the noisy image, $t(x)$ is the transmission rate and A is the atmospheric light. The larger the value of $t(x)$, the larger is the clarity of the object through the noise, and the larger the value of A, the whiter the added noise is. Three types of artificial noise were used to simulate the noise: uniform noise, linear-form noise, and exponential-form noise. In the uniform noise, $t(x)$ and A are a random number from 0.1 to 0.4 and 0.6 to 1, respectively. In the linear-form noise, from the center of the noise to the edge, $t(x)$ is linearly increased from 0.1 to 0.4 and A is linearly decayed from 1 to 0.6. In the exponential-form noise, from the center of the noise to the edge, $t(x)$ is exponentially increased from 0.1 to 0.4 and A is exponentially decayed from 1 to 0.6. A random number of artificial noise spots ranging from three to eight with 20% uniform noise, 40% linear-form noise, and 40% exponential noise were randomly added to clean images. Examples of images with artificial noise are shown in Figure 5. In total, 14400 artificial noisy images are used to train the image denoising model.

(Figure 5) Examples of artificial noisy images. 3 kinds of artificial noise, e.g., uniform, linear-from, and exponential-form noise are used to simulate real noise.

To evaluate the performance of the denoising model, pigs in 884 clean images were labeled to train the YOLOv5 model and 354 real noisy images were labeled to test it. The size of the images used in the pig detection module is 256×256.

3.2 Experimental Environment and setup

All the experiments were conducted on a computer with a Ubuntu 20.04 operating system, PyTorch deep learning library [10] 1.8, NVIDIA GeForce RTX 3080, Intel Core i7-6700K 4.0GHz and 32 GB of RAM.

The loss function used to train the generator of the denoising model is an ensemble loss function consisting of an adversarial loss, a content loss, and a perceptual loss. The loss function used to train the discriminator consists only of an adversarial loss. The denoising model was trained for 200 epochs with a batch size 16 and the learning rate was linearly decayed from 0.0002 to 0 during each epoch. Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ was used as the gradient descent optimizer. The pig detection model was also trained for 200 epochs with a batch size of 16 and the learning rate was linearly decayed from 0.001 to 0 during each epoch. Adam optimizer with $\beta_1 = 0.937$ and $\beta_2 = 0.999$ was used as an optimizer.

3.3 Results & Analysis

Precision, Recall, and Average Precision (AP) were used to evaluate the pig detection performance on noisy images and on images denoised through the preprocessing module. Table 1 shows the comparison results of both experiments.

<Table 1> Comparison of pig detection results on noisy images and denoised images

	Noisy images	Denoised images
Precision	0.741	0.787
Recall	0.66	0.724
AP	0.639	0.759

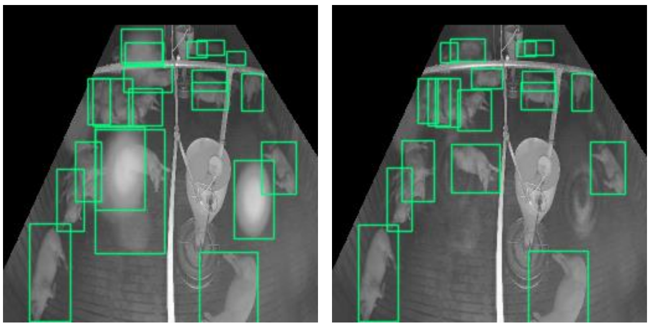
After applying the preprocessor, the precision, recall, and average precision improved greatly from 0.741, 0.66, 0.639 into 0.787, 0.724, 0.759, respectively. The results show that the proposed system can improve the robustness of the pig's automatic detection system in the real pig farms where the harsh environment always affects the pig detection performance.

The inference time of the preprocessor and pig detector is 1.9ms and 2.0ms, respectively. The inference time proves that the proposed system can be applied in real-time.

An example of an image denoised through the image denoising model is shown in Figure 6 and an example of the pig detection results on a noisy image and a denoised image are shown in Figure 7.

(a) Real noisy image (b) Denoised image

(Figure 6) Example of real noisy image and its corresponding denoised image produced by the image denoising model.



(a) Real noisy image (b) Denoised image

(Figure 7) Example of the detection results on real noise image and denoised image.

4. CONCLUSION

To improve the robustness of pig automatic monitoring in real pig farms where infrared cameras are always polluted due to the harsh environment, in this paper, we propose a real-time noise-robust infrared camera-based pig automatic monitoring system that adds a preprocessing step which transforms noisy images into denoised clean images before pig detection. The results show that adding the preprocessor improves greatly the performance of the pig detection and confirms the practicability of our proposed system. In the future, we will introduce an attention mechanism to the denoising model so it can put more focus on the noise area to produce clearer denoised images and improve the pig detection results.

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] KOSIS 국가통계포털, https://kosis.kr/index/index.do [accessed Sep. 16, 2021].

[2]농림축산식품부, https://mafra.go.kr/sites/mafra/index.do [accessed Sep. 16, 2021].

[3] J. Sa, Y. Choi, H. Lee, Y. Chung, D. Park, and J. Cho, "Fast Pig Detection with a Top-View Camera under Various Illumination Conditions," Symmetry, Vol. 11, 266, 2019.

[4] J. Seo, H. Ahn, D. Kim, S. Lee, Y. Chung, and D. Park, "EmbeddedPigDet—Fast and accurate pig detection for embedded board implementations," Applied Sciences, Vol. 10, 2878, 2020.

[5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, pp. 2341-2353, 2011.

[6] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," IEEE Transactions on Image Processing, pp. 5187-5198, 2016.

[7] P. Isola, J.Y. Zhu, T. Zhou, and A.A. Efros, "Image-to-image translation with conditional adversarial networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125-1134, 2017.

[8] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8202-8211, 2018.

[9] GitHub - ultralytics/yolov5: YOLOv5 in PyTorch > ONNX > CoreML > TFLite, https://github.com/ultralytics/yolov5 [accessed Sep. 16, 2021].

[10] PyTorch, https://pytorch.org [accessed Sep. 16, 2021].