Advances in Artificial Intelligence and Machine Learning; Research 4 (2) 2258-2275 Received 15-03-2024; Accepted 20-05-2024; Published 27-05-2024

Okeydoggy: A Mobile Application That Helps Understand Calming Signals of Dog Using MOBILENETV2 —Focusing on "Play Bowing" and "Licking Nose"

Myungjin Kang kangel429@nate.com
Department of Art & Technology
Sogang University
Seoul, South Korea

Rim Yu limlim_e@naver.com
Department of Art & Technology
Sogang University
Seoul, South Korea

Yongsoon Choi goodsoon96@gmail.com
Department of Art & Technology
Sogang University
Seoul, South Korea

Corresponding Author: Yongsoon Choi

Abstract

Domestic dogs display a range of postures, facial expressions, and other body language cues that can be used to evaluate their emotional and mental state; however, many do not understand the meaning of theses signals. In this study, we recognized calming signal of companion dogs using a mobile application powered by MobileNetV2. Specifically, we recognized calming signal such as "Play bowing", which indicates to reinitiate play after a pause, and "Licking nose", which indicates nervousness. We conducted a user survey with 44 dog owners to investigate if this mobile application provides positive user experiences, and received positive responses.

Keywords: Companion animal, Companion dog, Communication, Pet tech, Artificial intelligence, Deep learning, Image recognition

Citation: Myungjin Kang, et al. Okeydoggy: A Mobile Application That Helps Understand Calming Signals of Dog Using MOBILENETV2 —Focusing on "Play Bowing" and "Licking Nose". Advances in Artificial Intelligence and Machine Learning. 2024;4(2):130.

1. INTRODUCTION

1.1 Research Background

Dogs use calming signals, i.e., body language, as a means of communication. Rugaas (2006) indicated some behaviors displayed by domestic dogs were able to de-escalate or interrupt an aggressive encounter. Such signals were hypothesized by Rugaas to be able to prevent aggressive episodes to avoid conflicts. Rugaas labeled these signals as "calming signals." She argued that dogs use calming signals to calm themselves when they feel stressed or nervous, to communicate with other dogs and to get along with other dogs or people [1]. Mariti et al. (2017) conducted a scientific analysis on whether such signals can also calm other dogs in addition to communicating with them. Mariti et al. mentioned several examples of calming signals including Head turning, Turning away, and Play bowing [2].

The general method of recognizing canine behavior is to attach gyro or acceleration sensors to their bodies [3, 4]. Wearable devices equipped with sensors are therefore placed on dogs for behavior recognition. With recent developments in deep learning, however, several studies have been conducted on animal behavior recognition using image recognition and pose estimation rather than sensors.

We had initially planned to use the 2D or 3D pose feature for the present study. However, there is a lack of datasets for pose estimation in dogs, unlike humans. DeepLabCut [5] is a study that allows researchers to overcome the problem of insufficient data in animal studies such as this one. Specifically, DeepLabCut is a software package that stores the joint positions of an animal and uses such data to train a multi-layer artificial neural network for tracking animal movements. However, because dogs are quadrupedal, it was difficult to obtain accurate 2D coordinates for the hidden physical parts when a dog is facing the camera. In addition, DeepLabCut cannot estimate animals' 3D poses from a single view camera.

Due to occlusion, 2D pose estimation is not suitable for analyzing the behavior of quadrupedal animals, and given the lack of 3D datasets, it was judged that estimating 3D poses from a single view camera would be difficult. We therefore decided that image recognition is more appropriate for extracting features from canine images and recognizing their patterns. However, it would be difficult to collect a large number of calming signal images from Google Images. As such, this study attempted to perform frame-by-frame image recognition.

According to Luo et al. (2018), the learning of CNN requires a large amount of images and label each image and it takes a lot of time and effort to generate the training data [6]. However, it was difficult to collect these images using image crawling, which gathers image data from Google. Transfer learning is efficient in transferring the knowledge from a pre-trained model to another model (generally of higher complexity) to be trained using the new available data. Transfer learning eliminates the requirement of initiating a fresh learning scheme every time with different datasets.

1.2 Related Work

There are various examples of image recognition technologies that can help to analyze calming signals in dogs, such as "Play bowing" and "Licking nose." There are various technologies available for automated image analysis, including object detection, image classification, and image description generation, but the technologies that we have focused on in this research are Microsoft's What-Dog.net [7] and Microsoft's Azure Computer Vision API [8]. What-Dog.net is a service that applies a reverse image search technique. When a dog's picture is uploaded, it identifies the dog's breed, tells its temperament, and finds similar dogs. Azure Computer Vision API is also based on image recognition technology. By observing an image of a cat lying on a leather chair, it detects list of descriptions ordered from highest to lowest confidence. What is noticeable regarding the study is that the image recognition technology not only detected that the subject in the picture was a cat but also detected the cat's posture—"the cat is lying down." This implies that image recognition technology can be used to estimate not only "what the subject is" but also "what the subject's posture is."

We identified mobile applications for companions dog. There is a mobile application that can classify a dog's emotional state by recognizing the dog's voice with a deep learning algorithm. Another mobile application classifies the dog's emotional state through image classification and sends the information to the user.

Petpuls [9] is a collar-type wearable device worn by a dog, which is linked to the dog owner's mobile application. Its deep learning algorithm analyzes the physiological responses, such as the dog vocalization, respiration volume, and heart rate of the dog wearing Petpuls in real-time by combining with information, such as the breed, age, gender, and size of the dog. Based on this, it analyzes five emotional states of the dog: stable, happy, nervous, angry, and sad, and provides them as application notifications. Petpuls created a database by collecting voice sounds of approximately 10,000 dogs by breed and size for approximately three years. According to joint research conducted with the Music and Audio Research Group at Seoul National University, the emotion recognition accuracy exceeded 80%, which is expected to increase even more in the future as more data are accumulated.

However, according to the reviews on Petpuls at Naver's Smart Store [10] in South Korea, there was a drawback where the emotional state was difficult to determine when the dog was not barking because the emotional state was analyzed based on the dog's voice recognition.

A research team at the University of Melbourne used convolutional neural network (CNN) to develop Happy Pets [11], an application that classifies five emotional states: happy, angry, neutral, sad, and scared. Happy Pets learns to link facial expressions and emotions through thousands of image data, based on which the dog's emotion is detected. However, the emotions of the dog breeds that the Happy Pets can read are limited, and certain breeds may be classified as different breeds, reducing the accuracy. Considering the future plan of Happy Pets, which plans to improve the accuracy by including not only the dog's facial expression but also the body in the emotion analysis, it appears difficult to determine the emotion by relying solely on the dog's facial expression.

In particular, according to the reviews on Happy Pets at Apple's Appstore [12] in South Korea, the recognition results were different from the dogs' actual emotions. As mentioned above, we

identified a user opinion that “the emotional state is difficult to determine when the dog is not barking” in the case of Petpuls and that “it is difficult to determine the emotion by solely relying on the dog’s facial expression” in the case of Happy Pets. In this study, based on these user feedbacks, we aim to analyze calming signals to understand their emotions even when they are not barking. Rather than interpreting the potentially ambiguous changes in a dog’s facial expressions, they will focus on analyzing clear and distinct body language to determine the dog’s emotional state. Since dogs cannot speak, understanding their calming signals can be an effective way to communicate and interpret their emotions.

2. MATERIALS AND METHODS

Because the goal of this study is to develop a service application for mobile devices, we decided that MobileNets [13, 14, 15], which can be applied on mobile devices, was the most appropriate backbone for the present study. MobileNets are lightweight CNN architectures primarily designed for mobile and embedded device applications. Along with convolution layers for feature extraction, MobileNets employ depthwise separable convolution to significantly reduce the number of parameters in a neural network, thereby decreasing its size. MobileNet also uses two hyperparameters – namely the width = and resolution multipliers – in order to shrink the models [16]. According to Deng et al. (2009), a Neural Network (NN)-based detector typically uses a backbone network to extract basic features for detecting objects, and in most cases it is designed originally for image classification and pre-trained on ImageNet. We needed an image recognition method that uses transfer learning to allow us to achieve image classification even with a small number of images [17]. We recall the pre-trained MobileNetV2 model and proceed with the training again with the dataset we collected to suit the purpose of the study. When the training of the model is completed, the training model can be converted into the TensorFlow Lite [18] format for use on mobile devices. Finally, it is possible to develop a mobile application capable of image recognition by applying the extracted TensorFlow Lite to Android Studio.

2.1 System Architecture

FIGURE 1 shows the overall system structure of this application. First, we collect image datasets of calming signals of companion dogs, signs of wanting to play together, and feeling nervous to learn the model. Transfer learning is performed based on the MobileNetV2 model pre-trained with this data. Convert the trained model to Tensorflow Lite, and build Tensorflow Lite in Android Studio to create a mobile application that recognizes calming signals in real-time.

2.1.1 Dataset – selection of calming signal cues

In this study, images of calming signals of dogs, such as “Play bowing” and “Licking nose”, which signify that the dog wants to play together and feels nervous, respectively, are recognized using image recognition technology. As shown in TABLE 1, major calming signals of dogs defined by Rugaas [1] include Play bowing, Licking nose, and Splitting up.

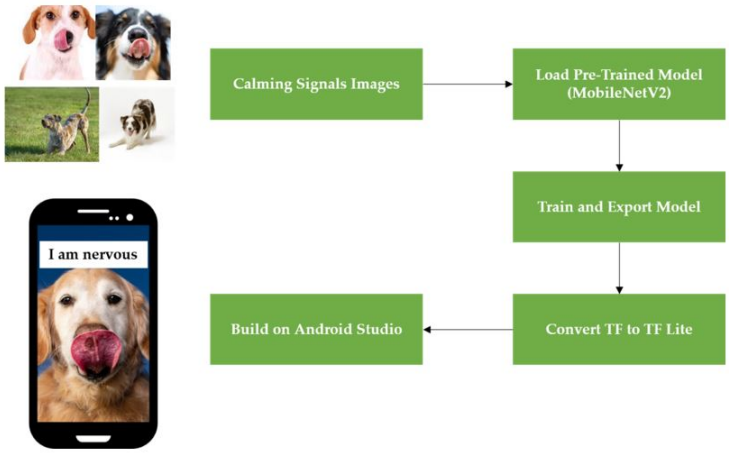

Figure 1: Overall system structure

According to Rugaas [1], bowing can be an invitation to play, particularly if the dog is jumping from side to side in playful manner. Furthermore, Licking nose can be seen in situations in which a dog might feel uncomfortable such as when its owner hugs it too tightly or speaks in an angry tone. Among the calming signals, Play bowing and Licking nose were defined as calming signals data because they showed a distinct motion difference based on the image recognition technology. In the future, we plan to expand classifiable calming signals gradually by enhancing image recognition technology.

Table 1: Classification of calming signals

Calming signal	Explanation
Licking nose	An act of quickly or slowly licking the nose with the dog's own tongue
Yawning	An act of yawning
Lifting paw	An act of slightly lifting one front leg
Freezing	An act of suddenly standing still without moving
Sitting down	An act of sitting down from the standing position
Lying down	An act of lying down from the standing or sitting position
Marking	An act of suddenly urinating
Play bowing	An act of raising the rump upward and stretching the legs straight forward
Wagging the tail	An act of wagging the tail
Head turning	An act of turning the head sideways or backwards or putting it in one direction
Turning away	An act of slightly turning the body sideways or completely turning away
Sniffing	An act of sniffing here and there on the floor or staying at a place for a while with the nose placed there
Curving	An act of suddenly walking in a curved form, such as a bow, when walking in a straight line
Splitting up	An act of cutting in when there are multiple dogs or people

Focusing on the calming signals of dogs, such as recognition technology "Play bowing" showing the intention of playing together and "Licking nose" expressing nervousness, we developed OkeyDoggy, a mobile application, which displays a message according to the calming signal. For example, it displays "Let's play" when the dog bows and "Nervous" when the dog licks the nose.

2.1.2 CNN network

Transfer learning is an important tool in deep learning to solve the problem of small datasets or insufficient training data [19]. It is the process of transferring the weights of a CNN model that has been trained on other large datasets such as ImageNet [17]. The pre-trained network is deprived of the classification layers and new empty classification layers are attached [20]. After that, the model is trained on a new dataset. Nowadays, a wide variety of CNN models have been introduced such a ResNet50 [21], AlexNet [22], InceptionV3 [23]. However, as much as these CNN models have been known for their performance, accuracy and their adaptivity to different datasets, they have a large number of layers and parameters which usually slow down the training and prediction operations. Newer models have been introduced to solve that problem, such as MobileNetV1 [13], MobileNetV2 [14] and MobileNetV3 [16]. The three versions of the MobileNet models has been improved ever since they were developed in 2017 [13]. The main purpose of the MobileNets was to implement a light CNN model on mobile devices with a reduced model size (<10 MB) and a reduced number of parameters.

In this study, we trained ResNet50, VGG-16, and MobileNetV2, and compared them in terms of loss and accuracy. All three models employ CNN structures, and have exhibited high performance in image recognition tasks. VGG-16 comprises 16 convolutional layers and a highly standardized architecture. Owing to its relative simplicity and small number of hyperparameters, it is a popular choice for extracting features from images using transfer learning [24]. ResNet50 is a 50-layer deep CNN with 48 convolutional layers, one max pooling layer, and one average pooling layer. ResNet links the nth layer input's directly to the (n+x)th layer, enabling the stacking of additional layers [25]. MobileNetV2 is a deep CNN with 53 layers, which functions as a powerful feature extractor that can detect and segment objects. It is built on an inverted residual structure with residual connections between bottleneck layers [26].

In this study, the model was trained with six classes as inputs to recognize the calming signals of dogs, such as "Play bowing" and "Licking nose," which have the meanings of wanting to play together and feel nervous, respectively: 751 images of "Play bowing" (Let's play), 848 images of "Licking nose" (Nervous), 1,049 images of "Standing up", a common posture of dog, 944 images of "Sitting down", 984 images of "Lying down" and 799 images of other cases where there is "No dog". Data were collected from Google Images, using combinations of the following search keywords: "Dog", "Play bowing", "Licking nose", "Standing up", "Sitting down", and "Lying down". However, because images that do not contain a dog may be misclassified, we created a "No dog" class to prevent this situation. The "No dog" class primarily encompasses images of spaces, such as typical cities and parks, instead of dogs. Because the collected images exhibited variable resolutions, all images were resized to 224 x 224 pixels prior to training. The images were collected randomly, without limitations according to dog breeds. Moreover, the collected images were mostly taken from angles directly facing the dogs during daytime when the subjects were visible.

In this study, only "Play bowing" and "Licking nose" were considered calming signals, and the app performed a binary emotional analysis in which owners were informed of their dogs' friendly emotions in the case of "Play bowing" and negative emotions in the case of "Licking nose." In addition, "Sitting down" and "Lying down" were included among the calming signals, but they were considered common postures because there is a high possibility that these two behaviors are dog behaviors that generally occur without any particular intention rather than being calming signals.

In addition, because behavior results were detected with very low probabilities even in cases where there were no dogs, images of general landscapes, people, etc. were classified as the "No dog" class to improve performance. The ratio of train and validation dataset was set as 8:2. We used the "splitfolders.ratio" specifying the true path of the folder containing the data that we want to split, the output true path location, and the ratio that we want to use to divide the data.

Since there is a limit to collecting a large amount of data corresponding to the calming signals of dogs, the problem was solved by augmenting the small number of data using the data augmentation technique. Data augmentation is a method of augmenting data by editing a picture into multiple pages by means of Flipping, Cropping, Rotation, etc. it is a technique used in computer vision and deep learning to artificially increase the size of a dataset by creating new, slightly modified versions of existing images.

In this study, the MobileNetV2 model was used as a pre-trained model for transfer learning. Tensorflow was used as this study's deep learning framework, Categorical Cross Entropy as its loss function, and RMSProp(Root Mean Square Propagation), as its optimizer. The RMSProp optimizer was used owing to its superior performance over the Adam optimizer. As hyper-parameters of the training model, the number of epochs was 100, the batch size was 32, and the learning rate was 0.0001. GeForce RTX 3090 with 64GB RAM was used as the graphical processing unit in all the experiments.

2.1.3 Mobile application

If the APK file is extracted by converting the trained model into the TensorFlow Lite format, and the mobile application is executed, it is possible to analyze the dog's behavior observed through the camera. APKs are a file format used for the installation and distribution of mobile apps.

The TensorFlow Lite Support Library provides a TensorImage class that can contain images to be input to the model. When images are collected in real time through the camera API, the library converts Bitmap format images into the ByteBuffer format, allowing them to be directly input to the model. The model classifies images by matching similar label values, and the prediction results are displayed on the mobile app screen immediately. FIGURE 2 shows that the application determined that a dog's observed behavior is "Play bowing" and displays a message for the pertinent calming signal, "Let's play."

The OkeyDoggy application has a function for showing the calming signal in the lower bar of the main screen when the entry screen changes to the camera screen. If the accuracy of the calming signal recognized on the camera screen is over 95%, the result of which calming signal was recognized is displayed. When the dog is viewed through the camera while the calming signals button is pressed, if a calming signal is recognized, a message expressing the dog's current state is displayed on the center screen.

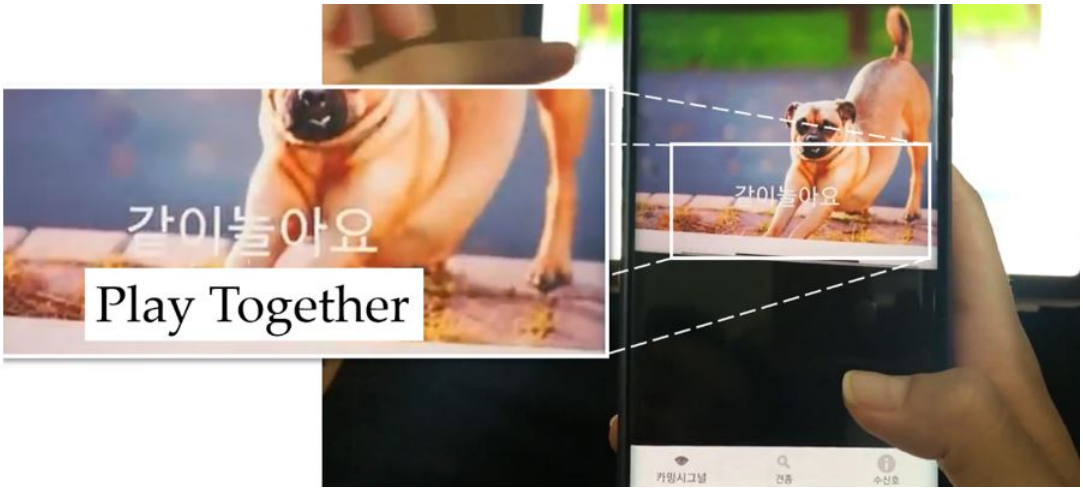

Figure 2: Final Prototype of OkeyDoggy

2.2 User Experience Methodology

In this study, we defined hypotheses H1 and H2 to determine if the application helped users to understand their dogs' calming signals and hypotheses H3 and H4 to determine if the application provided users with a positive experience. TABLE 2 shows research hypotheses and their corresponding sub-hypotheses.

Table 2: Research hypotheses

Hypothesis	Description
H1	The dog owner can understand the dog's calming signal information provided by the application.
H2	The dog owner can trust the dog's calming signal information provided by the application.
H3	This application is desirable.
H4	This application is credible.

To determine if the application helped users to understand their dogs' calming signals, we measured how much the users understood and trusted the calming signal information provided by the application. According to Sukhwal et al. [27], information understandability refers to readability and interpretability of information to the readers. According to Tseng et al. [28], information reliability refers to whether the information is felt reliable. We brought questionnaire questions from the understandability scale of Ramaprasad [29] for the information understandability and from the reliable scale of Brackett et al. [30] for the information reliability.

Morville (2005)'s honeycomb model is a tool that validate user experiences from a holistic perspective and includes measurement factors such as usable, useful, desirable, valuable, findable, accessible, and credible [31]. To determine if our mobile application provided users with a positive experience, we selected desirable and credible from among the honeycomb model's seven measurement factors to verify useability.

There are two questions for each evaluation item, with a five-point Likert scale per question, resulting in a total of 8 questions. In each question, the score was 1 point for "strongly disagree," 2 points for "disagree," 3 points for "neutral," 4 points for "agree," and 5 points for "strongly agree".

2.2.1 Statistical analysis

From May to June 2021, a user survey of 47 dog owners was conducted at parks in Seoul and Incheon, South Korea. During that period, the dog owners walking dogs in the parks executed the application, checked the screen configuration, and then executed the user tasks sequentially.

The experiment was conducted under highly restrictive conditions. Specifically, it was conducted during daytime, in parks with monotonous backgrounds that did not interfere with the recognition of dogs. Thus, dog owners performed the user tasks in an environment where the entire subject was clearly visible with minimal light interference. Owners were advised to point the camera toward their dogs from a frontal angle, directly facing the dog, to capture the entire subject, as the app is unable to identify the dog in the image if the entire subject is not captured. Owners were then given smartphones equipped with the app to perform user tasks. The dogs took various poses on a green area.

Subsequently, they answered the survey questions, and later, we conducted in-depth user interviews. For the user survey, we obtained permission from all the dog owners, and the entire process was recorded with audio and video. SPSS statistical software (IBM SPSS Statistics) was used to statistically analyze the survey.

We surveyed 47 test subjects who had companion dogs; however, determining the data for three persons were noises based on the data analysis results, we analyzed the data based on the responses of 44 final test subjects. As shown in TABLE 3, there were 14 males (31.8%) and 30 females (68.2%) among the 44 test subjects, and in terms of age, two persons were in their 10s (4.5%), 21 persons in their 20s (47.7%), 14 persons in their 30s (31.8%), and seven persons in their 50s or older (16%). Because 22 persons (50.0%) had companion dogs for less than or equal to five years, and 15 persons (34.1%) for 10 to 15 years, indicating that many test subjects did have a long experience of raising dogs, it is predicted that the companion dogs have increased sharply in recent years. Furthermore, 14 people (31.8%) responded that they knew calming signals, and 30 people (68.2%) responded that they did not know calming signals, indicating that the majority of the test subjects did not know calming signals.

Nunnally (1978) argued that if the Cronbach's α value was greater than or equal to 0.60 in the exploratory research field, the data were reliable [32]. The survey questions that were used in this paper had a Cronbach's α value of 0.882, and because the Cronbach's α value was higher than 0.60, it was determined that every evaluation item was a measurement indicator with inner consistency. Also, because the "Cronbach's alpha if item deleted" for all questions was between 0.820 and 0.874 as shown in TABLE 4, it can be seen that there were no questions that harmed reliability in particular.

Table 3: Demographics

	Category	N	%	Total
Gender	Male	14	31.8	44 persons
	Female	30	68.2	
Age	10s	2	4.5	44 persons
	20s	21	47.7	
	30s	14	31.8	
	40s	0	0.0	
	50s or older	7	16.0	
Period of raising companion dog(s)	5 years or less	22	50.0	44 persons
	5 to 10 years	4	9.0	
	10 to 15 years	15	34.1	
	15 years or longer	3	6.8	
I know calming signals	Yes	14	31.8	44 persons
	No	30	68.2	

Table 4: Reliability analysis

Category	Cronbach's Alpha if Item Deleted
Information Understandability	0.874
Information Reliability	0.820
Desirable	0.846
Credible	0.848

3. RESULTS

In this study, we compared the performance of three models: MobileNet2, ResNet50, and VGG-16. Among them, MobileNet2 exhibited the highest accuracy. Subsequently, we implemented a mobile application using MobileNet2 and evaluated the final user experience using this application. We assessed the accuracy for each class, the effects of data augmentation, and compared the accuracy and loss of the three models. Furthermore, we conducted descriptive statistics and correlation analysis for various user experience factors. Finally, based on depth interviews, we completed an affinity diagram.

3.1 System Evaluation

TABLE 5 shows the accuracy for each class: "Licking nose (Nervous)" has the highest accuracy at 98%, and "Standing up" has the lowest accuracy at 90%. The confusion matrix shown in FIGURE 3, refers to a table that compares the prediction and observation to assess the trained model's performance. In the case of "Licking nose (Nervous)", which showed the highest accuracy, the prediction matched the observation in 125 of 128 images, and it was incorrectly predicted as "Lying down" in three images. In the case of "Standing up," which showed the lowest accuracy, the prediction matched the observation in 142 of 158 images, and it was incorrectly predicted as "Licking nose

(Nervous)" in two images, "Other" in one image, "Sitting down" in nine images, "Lying down" in three images, and "Play bowing (Let's play)" in one image.

Table 5: Accuracy per class

CLASS	ACCURACY	SAMPLES
Licking nose (Nervous)	0.98	128
No dog	0.97	120
Sitting down	0.92	142
Lying down	0.92	148
Standing up	0.90	158
Play bowing (Let's play)	0.96	113

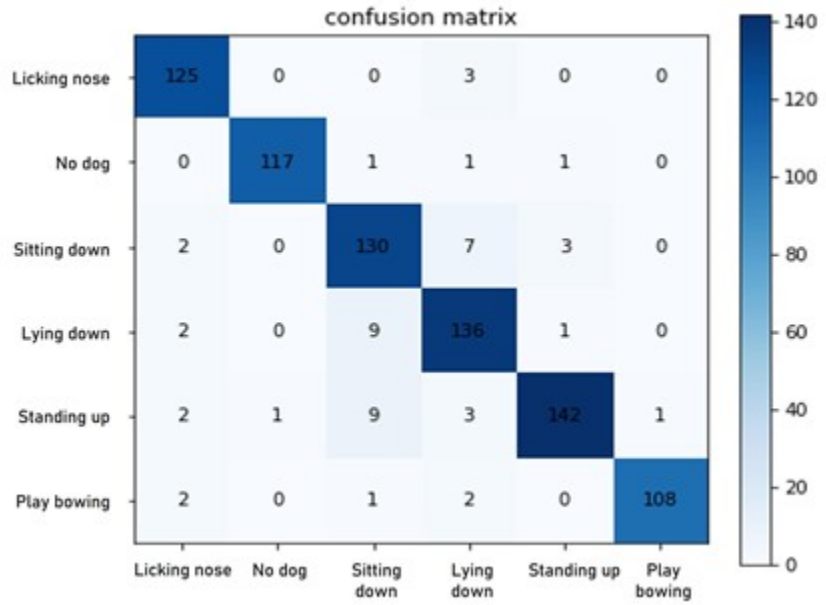

Figure 3: Confusion matrix

To find out how much the accuracy of the model improves when the data is augmented with the data augmentation technique, the accuracy and loss are compared before and after data augmentation. FIGURE 4(a), when training the model without augmenting the data, the train accuracy was 97.21%, the train loss was 0.11, the valid accuracy was 85.56%, and the valid loss was 0.2644. As shown in 7(b), when the model is trained after augmenting the data, the train accuracy is 99.30%, the train loss is 0.0517, the valid accuracy is 97.2%, and the valid loss is 0.0957, it can be seen that all values have improved compared to before.

We further tested the ResNet50 model and the VGG-16 model under the same conditions to verify the usefulness of the MobileNetV2 model. In FIGURE 5, we can see the results of comparing the accuracy and loss of each model on the validation dataset. When the model is trained with augmented data under the same conditions, the accuracy of the ResNet50 model is 68.4% and the loss is 0.5352, which shows the lowest performance among the three models. In addition, the

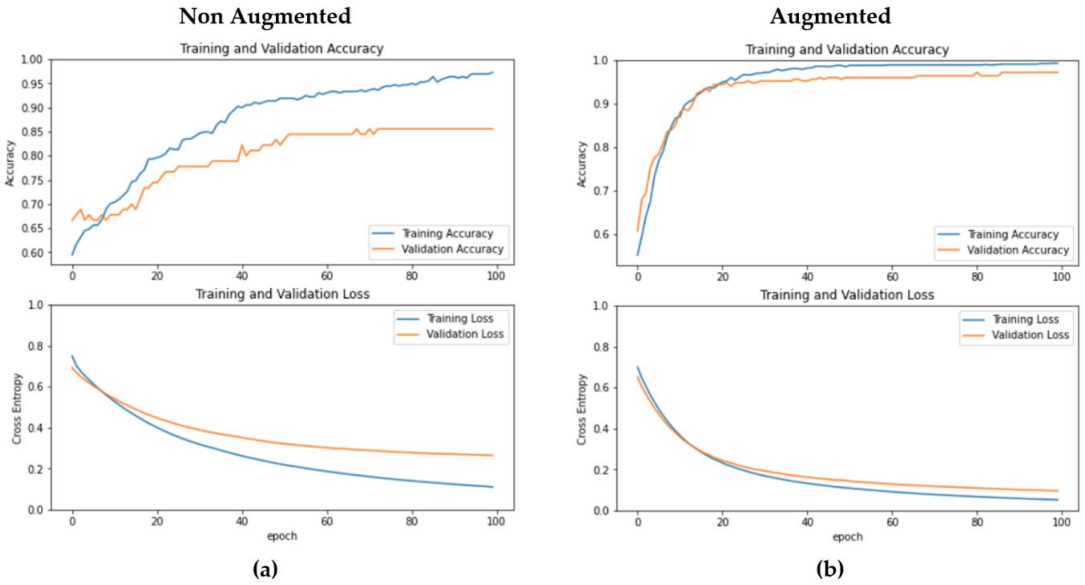

Figure 4: Effect of data augmentation

accuracy of the VGG-16 model was 89.2% and the loss was 0.3137, showing better performance than the ResNet50 model, but not reaching the performance of the MobileNetV2 Model.

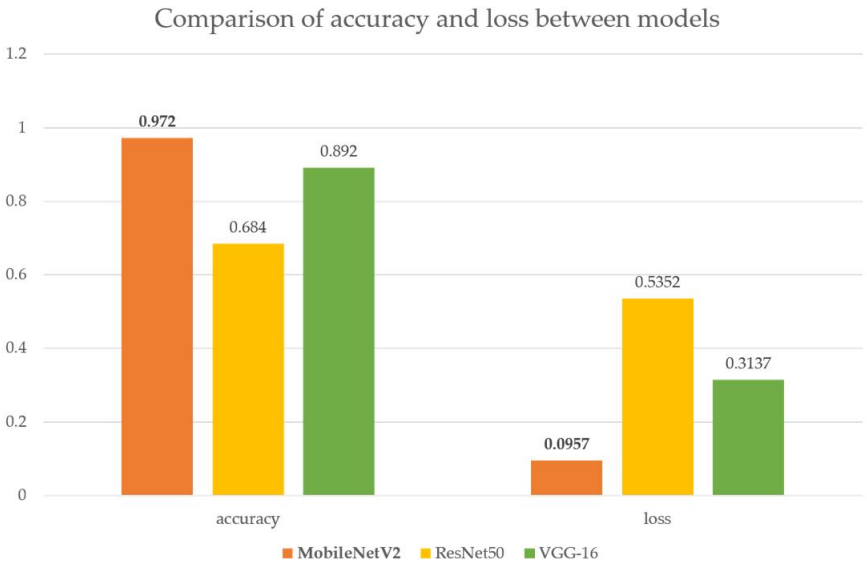

Figure 5: Comparison of accuracy and loss between models

3.2 Statistics Analysis

TABLE 6 shows the results of analyzing descriptive statistics to validate the research hypothesis. The results for "Information Understandability (M=4.5682, SD=.50106)" and "Information Reliability (M=4.3977, SD=.65218)" indicate that the overall levels for user experience factors are positioned between "Neutral" (3 points) and "Strongly agree" (5 points). Therefore, based on the determination that the "Information Understandability" and "Information Reliability" are at positive levels, the two detailed research hypotheses are accepted, and the app exhibits usefulness in allowing owners to understand their dogs' calming signals. Since "Information Reliability" is lower than "Information Understandability", it appears that the test subjects do not always trust the provided calming signal information. When analysis was performed based on in-depth interviews, it was observed that because there was no existing application comparable to this application, making it difficult to compare, the users had some difficulty in trusting the calming signal information provided by the application.

In addition, the "Desirable (M=4.4886, SD=.59542)" and "Credible (M=4.3523, SD=.75928)" results were judged to be positive levels, and we accepted the hypothesis that the app provides users with a positive experience in regards to the user experience evaluations of the app. However, the results of analyzing the in-depth interviews showed that "Credible" had a large standard deviation. This implies that the app worked well when it was used for short periods of time, but it became unstable when it was used for long periods.

Table 6: Descriptive statistics analysis

	N	Min.	Max.	M	SD
Information Understandability	44	3.5	5.0	4.568	.501
Information Reliability	44	3.0	5.0	4.398	.652
Desirable	44	3.0	5.0	4.489	.595
Credible	44	2.0	5.0	4.352	.759

3.3 Correlation Analysis

The Pearson correlation coefficients between "Information Understandability", "Information Reliability", and the honeycomb model's two measurement factors are presented in TABLE 7 (with the significance level set to 0.01).

When the correlations were analyzed, we found that "Information Understandability" and "Information Reliability" had positive (+) correlations with "Desirable", and "Credible".

The Pearson correlation coefficients between "Information Understandability," "Information Reliability," "Desirable," and "Credible" ranged from 0.529 to 0.697, confirming that they had fairly high correlation. This suggests that the more easily users were able to check their dogs' calming signals with the app, the better their experience of the app. Specifically, because "Information Reliability" and "Credible" had the highest correlation at 0.697, it can be seen that the more the owners felt they could rely on the calming signal information provided by the app, the more the app's credibility increased.

Table 7: Result of correlation analysis

	Information understandability	Information reliability	Desirable	Credible
Information understandability	1			
Information reliability	.751**	1		
Desirable	.529**	.686**	1	
Credible	.577**	.697**	.742**	1

*$p < .05$, **$p < .01$

3.4 Depth Interview

We conducted depth interviews with some test subjects to obtain qualitative evaluations of this application. We then used an affinity diagram to compare and analyze the opinions obtained through the in-depth interviews. Known as the KJ method, the Affinity Diagram was developed by Kawakita Jirou in 1953. After field interviews, all the participant notes are collated and audio or visual files are converted into transcripts. Keywords, phrases, or sentences relevant to the activity theme are recorded on cards or notes, and then sorted into groups [33]. We used the affinity diagram to categorize the opinions from the in-depth interviews into three categories, i.e., feeling, need, and actual situation, listed in FIGURE 6.

4. DISCUSSION

In this study, we developed an application that can recognize the dog's state through image analysis by training the model with a dataset of images corresponding to "Play bowing" and "Licking nose," which, in the calming signals gesture language used by dogs, convey the meanings of "Let's play" and "Nervous," respectively.

Through a system evaluation, we investigated the accuracy of the image analysis in the proposed application. We demonstrated the effects of data augmentation during network training by creating a graph of training, validation accuracy, and loss before and after applying data augmentation, and we compared the performance of a MobileNetV2 model, ResNet50 model, and VGG-16 model to show that the MobileNetV2 model had the best performance in this study.

Moreover, we developed a mobile application based on the MobileNetV2 model, capable of real-time analysis of calming signals in dogs. We conducted descriptive statistical analysis to assess the application's utility in understanding dogs' calming signals and its overall user experience. Furthermore, correlation analysis was performed to investigate any potential linear relationship between understanding calming signals and application usability.

We conducted Likert scale-based quantitative user experience evaluations based on the two evaluation items "Information Understandability" and "Information Reliability" for usefulness in understanding calming signals and the honeycomb model's two items for application usability. After

Figure 6: Affinity diagram

the experiment, we also conducted qualitative evaluations using interviews. Based on these, we confirmed that the designed application is actually useful for understanding calming signals and provides positive experiences to the users. In particular, we confirmed that the more the users felt they could rely on the calming signal information provided by the app, the more the app's credibility increased.

In addition, because this study selected the two least similar calming signals when performing image recognition, the app's accuracy may suffer when trying to classify different calming signals.

We compiled a dataset related to calming signals from Google Images, which was highly challenging as these signals are not generally recognizable. In particular, finding many images of calming signals for a certain dog breed was difficult. Instead, we collected these images independent of breed. However, because there are differences in the appearance of each dog breed, we believe more meaningful results could be obtained if the dataset and experiment were restricted to a single breed.

The experiment for the usability survey was conducted under very limiting conditions: during the daytime, and in parks with green areas. Consequently, the recognition rate may vary significantly in general circumstances. Specifically, the recognition rate decreased when the camera did not fully capture the canine subject. In addition, it seems that the recognition rate was good because training was performed with classes generally consisting of images captured in daytime. The experiment itself was also conducted during daytime. The recognition rate may drop at nighttime or indoors

when the subject is less visible. Moreover, the recognition rate was generally suboptimal when the camera was pointed toward the dog from the left, right, and rear, rather than the front.

While it is possible to achieve high accuracy using large and heavy models, these models present challenges when implemented in mobile or embedded devices. Instead, we used the lightweight MobileNetV2 to verify the real-time recognition of calming signals for dogs in a mobile app. We also demonstrated how this image recognition technology can be used in practice. However, many additional variables must be considered in real environments, including camera angle and light intensity. We verified that these variables significantly affect recognition results. Furthermore, we believe that the results would have been very different if the experiment had been conducted in a space with many objects, or at nighttime. Therefore, further research is needed to ensure that the app can be used immediately in practical settings without experimental constraints.

References

[1] Rugaas T. On Talking Terms With Dogs: Calming Signals. Dogwise publishing. 2006.

[2] Mariti C, Falaschi C, Zilocchi M, Fatjó J, Sighieri C, et al. Analysis of the Intraspecific Visual Communication in the Domestic Dog (Canis Familiaris): A Pilot Study on the Case of Calming Signals. J Vet Behav. 2017;18:49-55.

[3] Vehkaoja A, Somppi S, Törnqvist H, Valldeoriola Cardó A, Kumpulainen P, et. al. Description of Movement Sensor Dataset for Dog Behavior Classification. Data Brief. 2022;40:107822.

[4] Kumpulainen P, Valldeoriola A, Somppi S, Törnqvist H, Väätäjä H, et. al. Dog Activity Classification With Movement Sensor Placed on the Collar. In: Proceedings of the fifth international conference on animal-computer in-teraction; 2018:1-6.

[5] Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, et al. Deeplabcut: Markerless Pose Estimation of User-Defined Body Parts With Deep Learning. Nat Neurosci. 2018;21(9):1281-1289.

[6] Luo C, Li X, Wang L, He J, Li D, Zhou J. How Does the Data Set Affect CNN-Based Image Classification Performance? In: 5th International Conference on systems and informatics (ICSAI).2018;2018:361-366.

[7] https://www.bing.com/visualsearch/Microsoft/WhatDog#

[8] https://azure.microsoft.com/ko-kr/services/cognitive-services/computer-vision/#overview

[9] https://www.petpuls.net/petter

[10] https://smartstore.naver.com/

[11] https://apkplz.net/app/au.edu.unimelb.eresearch.happypets

[12] https://apps.apple.com/kr/app/happy-pets/id1515202735

[13] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, et. al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. 2017.

[14] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MOBILENETV2: Inverted Residuals and Linear Bottlenecks. 2019.

[15] Bhuiyan MA, Abdullah HM, Arman SE, Saminur Rahman SS, Al Mahmud K. Banana Squeeze Net: A Very Fast, Lightweight Convolutional Neural Network for the Diagnosis of Three Prominent Banana Leaf Diseases. Smart Agric Technol. 2023;4:100214.

[16] Qian S, Ning C, Hu Y. MOBILENETV3 for Image Classification. In: Proceedings of the 2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE), Nanchang, China. 2021:490-497.

[17] Deng J, Dong W, Socher R, Li LJ, J, Li K. Imagenet: A Large-Scale Hierarchical Image Database. In: IEEE conference on computer vision and pattern Recognition. 2009;2009:248-255.

[18] https://www.tensorflow.org/lite

[19] Tan C, Sun F, Kong T, Zhang W, Yang C, et. al. A Survey on Deep Transfer Learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN.Springer, Cham. 2018:270-279.

[20] Jeon Y, Choi Y, Park J, Yi S, Cho D, et al. Sample-Based Regularization: A Transfer Learning Strategy Toward Better Generalization. 2020. Arxiv Preprint: https://arxiv.org/pdf/2007.05181

[21] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.

[22] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. Advances in neural information processing systems. New York: Curran Associates, Inc. 2012;25:1097-1105.

[23] Dongmei Z, Ke W, Hongbo G, Peng W, Chao W, Shaofeng P. Classification and Identification of Citrus Pests Based on INCEPTIONV3 Convolutional Neural Network and Migration Learning. In: Proceedings of the 2020 international conference on Internet of things and intelligent applications (ITIA), Zhenjiang, China. 2020:1-7.

[24] https://iq.opengenus.org/vgg16

[25] Jahromi S, Buch-Cardona P, Avots E, Nasrollahi K, Escalera S, et al. Privacy-Constrained Biometric System for Non-cooperative Users. Entropy. 2019;21:1033.

[26] Seidaliyeva U, Akhmetov D, Ilipbayeva L, Matson ET. Real-Time and Accurate Drone Detection in a Video With a Static Background. Sensors. 2020;20:3856.

[27] Sukhwal A, Mathur A. Antecedents to Customer Acceptance of Information in E-word of Mouth. NMIMS Manag Rev. 2017;34:58-72.

[28] Tseng S, Fogg BJ. Credibility and Computing Technology. Commun ACM. 1999;42:39-44.

[29] Ramaprasad J. Informational Graphics in Newspapers. Newspaper Res J. 1991;12:92-103.

[30] Brackett LK, Carr BN. Cyberspace Advertising vs. Other Media: Consumer vs. Mature Student Attitudes. J Ad-Vertising Res. 2001;41:23-32.

[31] Morville P. Ambient Findability: What We Find Changes Who We Become. O'Reilly Media, Inc.. 2005.

[32] Nunnally JC. Psychometric Theory. 2nd ed. New York: McGraw-Hill. 1978.

[33] Duh HB, Lee JJ, Rau PL, Chen MQ. The Management Model Development of User Experience Design in Organization. In: International Conference on Cross-Cultural Design. Cham: Springer. 2016:163-172.