2021 6th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 02-04, 2021

Combined Approach of Supervised and Un-supervised learning for Dog Face Recognition

D.T. Weerasekara
Faculty of Computing,
Sri Lanka Institute of Information Technology
Sri Lanka
it17010122@my.sliit.lk

M.P.A.W. Gamage
Faculty of Computing,
Sri Lanka Institute of Information Technology
Sri Lanka
anjalie.g@sliit.lk

K.S.A.F. Kulasooriya
Faculty of Computing,
Sri Lanka Institute of Information Technology
Sri Lanka
it17011730@my.sliit.lk

***Abstract*- One would be surprised to hear the lost dog rates around the world. Even though it is something that one doesn't ponder a lot about, lost dogs are a problem that most dog owners fear. Dogs provide humans with companionship, protection, and unconditional love, and to the dogs; their whole world revolves around their owner and their family members. Therefore, when a pet dog goes missing, not only the dog owner but also the pet dog is affected. Unfortunately, in Sri Lanka, a lost dog being found is a very rare occurrence. A reason for this can be pointed out as the lack of an easily-accessible, public platform for lost dogs. In this research project, a solution to this problem has been implemented using image processing. This research study is about image classification and recognition using the Convolutional Neural Network (CNN) or also known as Shift Invariant or Space Invariant Artificial Neural Network (SIANN) by using TensorFlow framework as well as Keras library. The VGG16 model was customized for being used feature extraction. The implementation was a combination of both Machine Learning and Deep Learning. The platform to upload the found dog is also a continuous and inter-related subcomponent that provides a happy and healthy life for stray dogs too. That idea is providing them a higher chance to find a safe place to survive and also a home where they will be loved. The results are discussed in terms of the accuracy of the image recognition and classification in percentage. Each group of dogs get around 90% accuracy or above.**

Index Terms- Feature extraction, Dog matching, Deep Learning, Image Classification, TensorFlow, Keras.

I. INTRODUCTION

Dogs are considered to be a common component of most of the families in Sri Lanka and as well as many other countries. Having a pet dog is considered to be beneficial for humans in various ways. The main reason for having a pet dog in a house is for safety. Furthermore, owning a dog has many health benefits too. Regular walking and playing with a dog can decrease blood pressure, cholesterol levels, and triglyceride levels. Mainly pets can help manage loneliness and depression by giving us companionship. So, dogs play a main role in a house.When looking at society, it is very clear that pets are lost from the home regularly. But it is unclear how many of those dogs are returned and reunited with their family. Many dogs return to their home on their own or returned home by the neighbors mostly. Others may end up being directed towards a shelter until they are reunited with the owners. But some dogs might not be lucky enough to find their way to a safe place and be astray. Previous researches show that it is an important goal to be achieved to find a way to reunite the owners with their lost pets and it worth the effort. Providing security and a safe environment for a dog without chaining a dog is a very hard task.

At present, Image processing is a trending topic in the technological field. It diversely grows to conquer different parts of the other industries such as automobiles, healthcare, gaming, and e-commerce. The most well-known example is Facebook. Facebook is now capable of identifying one's face only with the knowledge of previously tagged images and your profile picture. No longer the technology will beat up the ability of human image classification power.

Deep Learning (DL)is the main dominant approach for this technology. DL is the ability of a computer to think like a human without the help of a human, which falls under Artificial Intelligence (AI). So, it is advisable to train the system with hundreds and thousands of data in order to improve the training sessions, so that the results will be efficient and fast. Machine Learning (ML) is also frequently applied in Image classification, But ML alone cannot be applied in Image classification, since it needs some improvement. This is a combined approach using both DL and ML. Computer vision comes from modeling image processing using the techniques of ML. Computer vision applies ML to recognize patterns for the interpretation of images[1].

In this paper, a Convolutional Neural Network(CNN) based on TensorFlow is used with python as the programming language. Around a thousand input data are used in this project. The accuracy of each pre-trained model will be studied and compared for different categories of dogs.

II. RELATED WORK

Computer vision is not concentrated only on Image or Video classification. This paper mostly discuss the technology behind the Image Classification using DL. DL's history runs to more than 20 years back. It can be proved by the LeNet[2] introduced during 1998, which can be set as the foundation of the present image classification using CNN. This network was able to achieve 99.05% accuracy for 20 epochs training. After a decade later in 2012, Alex et al.[3] trained a large deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet contest. This was able to achieve the least error rate of 15.3% which is comparatively very less than the second least which was 26.2%. To accomplish their target, they have inherited the multi-layer CNN idea from the LeNet but increased the size of the CNN bigger. The input of the AlexNet was able to reach 224 x 224, while LeNet was able to input only 32 x 32.

2021 6th International Conference for Convergence in Technology (I2CT) | 978-1-7281-8876-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I2CT51068.2021.9418175

In [4], Neural Network Architecture was studied as a method of image classification. Here the system slowly improves the MNIST models. MNIST is an open-source database that can be used as training sets. Furthermore in [5], image classification is discussed based on CNN models. There are several CNN models developed upto date. Here, the training was done balancing both face-containing images and non-face images. This research achieves an 81.6% detection rate with only 6 false positives on the Face Detection Data set.

Even though ML cannot be considered as a good approach for Image Classification, the research done by [6] has used Decision Tree (DT) as the technique for image classification. This method is very simple but identified as very efficient too. The approach here is a hierarchical classifier. This classifier allows rejecting classes on the middle layers. Similarly, [7]has used an ML approach which is a Support Vector Machine for image processing. Here the results showed that the images have higher resolution in terms of effectiveness in regularity.

III. METHODOLOGY

Image Processing to identify or recognize a human face is the most often discussed at different levels. A combination of DL and ML approach is taken in order to accomplish this outcome. Figure 1 shows, the overview of the steps taken to achieve this particular task. This scenario can be divided mainly into 3 phases.

A. Data Collection and Data Preprocessing

At this time of research, there was no open-source dataset available to be used in this implementation. Therefore, a dataset was built by the images available in google and from other websites such as Flickr[8] and adoption sites such as Animal Happy end and Tiko. We were able to get around 2500 pictures of 320 dogs. For this purpose, dogs with at least 5 images were selected so that the accuracy can be increased.

The dataset collected to be used in this model is comparatively small so that the accuracy that will be given will be not good enough to predict a reliable outcome. In order to increase the accuracy of the model, data augmentation was done. Data Augmentation is a technique to increase the diversity of a training dataset, by applying random but realistic transformations to the image[9]. The transformations, that are being used in this implementation are rescaling, rotation, height shift, width shift, horizontal flip, validation split, and zooming.

Initially, even though there were 1000s of images retrieved they were grouped and only 5 groups were considered and used in testing the model. Each group contained images of different types of dogs. Each type of dog contained hundreds of images with different sizes, angles, and different colors. The total number of all these dogs is 1350 and the division is shown in the Table 1.

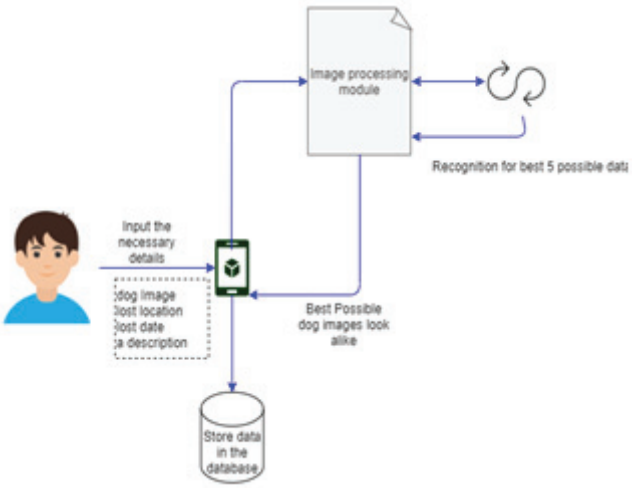

Fig. 1. System Overview Diagram. This shows how the complete approach is designed

TABLE I. NUMBER OF IMAGES ACCORDING TO THE TYPE OF DOGS

No	Types of dogs	No of Images
1	Class 0	295
2	Class 1	230
3	Class 2	263
4	Class 3	290
5	Class 4	227
	Total	1305

Fig. 2. Images of the "Class 4" dogs

Fig. 3. Images of the "Class 2" dogs

B. Training the dataset

To train the data properly, the dataset was split into two parts; training set and testing set. The testing data set can be divided into 2 other sets;

a. Closed-set
b. Open-set

Closed-set is a set of images that are considered to be unknown images of the known dogs. Simply explaining those are the images seen by the model during the training stage. The open set contains a mixture of pictures of unknown dogs. The system sees these images for the first time. These are harder problems for the system since the system needs to look into the problem closer.

Sometimes there will be overfitting occurring when the data is trained for more than the required amount of time. To avoid this overfitting the dataset was needed to be increased.

C. Implementation of Convoluted Neural Network (CNN)

As mentioned above, there are images of 5 types of dogs in different categories. This undergoes training with multiple layers mentioned under "Model conversion". This convolution process is done by the VGG16 model.

1) General Flow of the model

The figure shown above is the general scenario behind the model. As explained above the data set is divided into two parts and feature extraction is performed. This will be further explained in the next part of the document. Finally, the model is trained and the trained model is used in the classification and predicting the final algorithm.

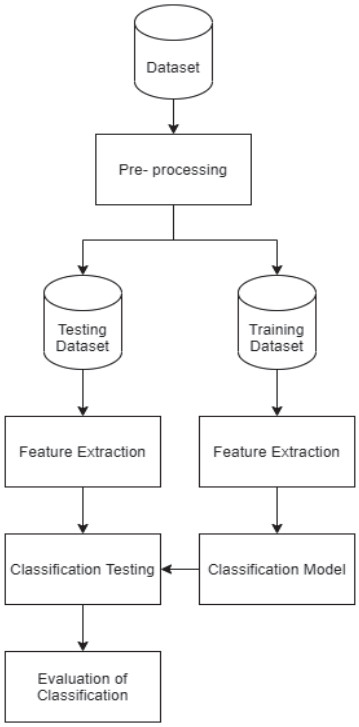

Fig. 4. The general pipeline of the model

2) Model Conversion

This system needs to possess a higher accuracy because this is surely going to provide you with real-time answers. So, if the system prediction is going to predict a distinguishable different dog image, that can be identified by the user. So, to overcome this issue the model should be trained well. Therefore, rather than training our own model, it is much better to use the weights of a pre-trained model. VGG16 is a model that can be used in such situations. VGG16 is a CNN model proposed by K. Simonyan and A.Zisserman. This model is confirmed to achieve an accuracy of 92.7% in the ImageNet dataset.

There are altogether 16 layers; convolution and max-pooling. This VGG-16 model can be downloaded from the Keras application. By using the sequential API, all the layers are passed to our model except for the last layer. The last layer is excluded because that particular layer consists of 1000 output labels/class. Since we do not need these 1000 classes it is opted out and add some layers of our model and train our model. Simply we freeze the first 15 layers and remove the last or final layer and add the layers of our model at the end. The concept behind this is called **Transfer Learning for fine-tuning** and also **feature extraction**

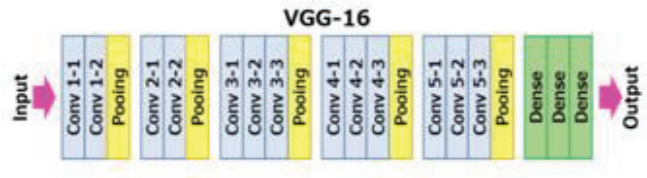

Fig. 5. VGG-16

Layer (type)	Output Shape	Param #
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312

Fig. 6. Using the 15 layers of VGG16 model

dense_1 (Dense)	(None, 512)	2097664
dense_2 (Dense)	(None, 128)	65664
dense_3 (Dense)	(None, 128)	16512
dense_4 (Dense)	(None, 5)	645

Fig. 7. Passing the system layers to train the model

The next step is to train the above built customized model. To train this model Adam optimizer is used in Keras. Even though there are many other optimizers, Adam is considered to be the best out of them. The loss function used is categorical crossentropy. That is because there are more than two classes. Then the model is saved. If the model is already saved, the model should be loaded and compiled. This compilation should be done definitely.

The model created above is not the exact model needed for the system. The features in the above three layers (dense1, dense2, dense 3) will be taken separately except for the last layer. The reason is the last layer contains only 5 neurons, but the above layers have 128 and 512 neurons. The strategy behind this mechanism is we pass the input images to the VGG16 model and create a set of intermediate high-level features. By using this feature vector, the neighbors are predicted.

Finally, using the **NearestNeighbors** method in sklearn, the neighbors are set to fit with the test features. Then using **neighbor.kneighbors** the indexes of the nearest neighbor can be extracted.

As explained above, the research component is being developed using a combination of DL and ML. This system is also a combination of unsupervised and supervised learning too. The flow is explained further in the given diagram.

Fig. 8. Flow of used Technological background

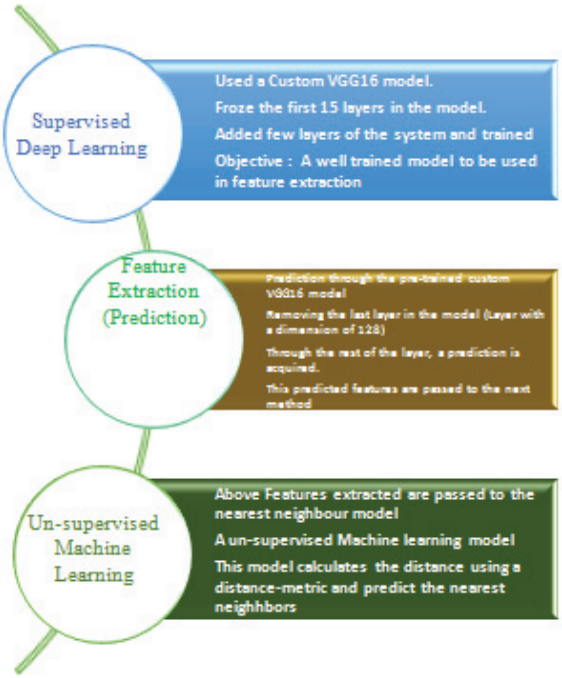

Fig. 9. Technological Overview explained

IV. Results And Findings

The above-implemented model is tested initially by testing with 20 epochs and later using 30 epochs, the distance achieved at both the instances is as follows. When the model is trained with 20 epochs, the results for the input image are as follows. Even though the same breed as predicted, there was to be a distinguishable difference between the predicted images.

```
Neighbour image 1 Distance : 8.969100758732482
Neighbour image 2 Distance : 10.644658995194536
Neighbour image 3 Distance : 11.173884796336758
Neighbour image 4 Distance : 12.308127563273217
Neighbour image 5 Distance : 13.715187430064944
```

Fig. 10. Distance of the images after 20 epochs

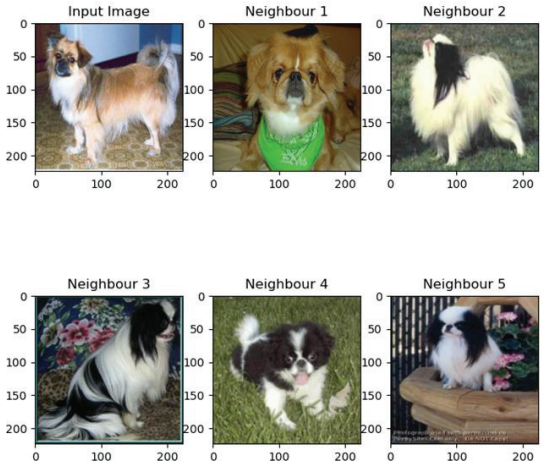

Fig. 11. The output images after 20 epochs

Next, to increase the accuracy of the model, this was trained for 30 epochs and the model results obtained were as following

```
Input image label : 1
Neighbour image 1   , Class : 1   ,  Distance : 1.9695574345280712e-25
Neighbour image 2   , Class : 1   ,  Distance : 2.247902979741616e-25
Neighbour image 3   , Class : 1   ,  Distance : 2.2689215609997456e-25
Neighbour image 4   , Class : 1   ,  Distance : 2.269309571485266e-25
Neighbour image 5   , Class : 1   ,  Distance : 2.269404634943807e-25
```

Fig. 12. Distance of images after the 30 epochs

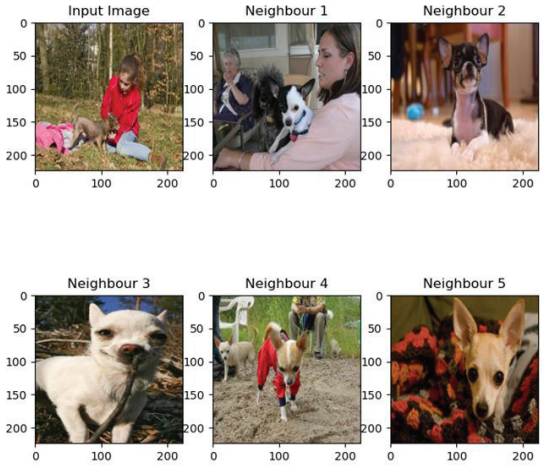

Fig. 13. The output images after 30 epochs

The differences between the distances and the classes can be elaborated as follows.

The accuracy and the loss of the training and the validation set are shown in the figures given below. This would be seen continuously if the number of epochs considered is increased. But a considerable accuracy is achieved at a low number of epochs it was stopped there.

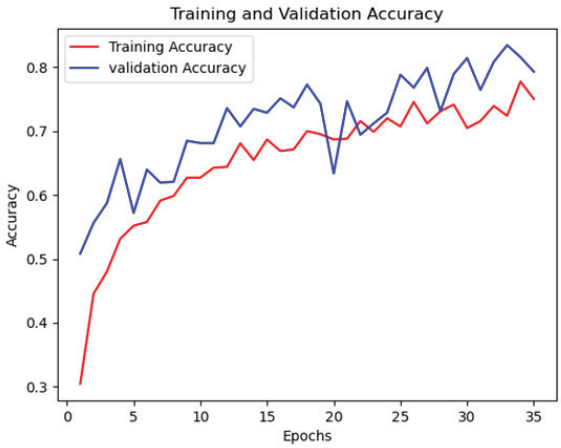

Fig. 14. Training and Validation Accuracy

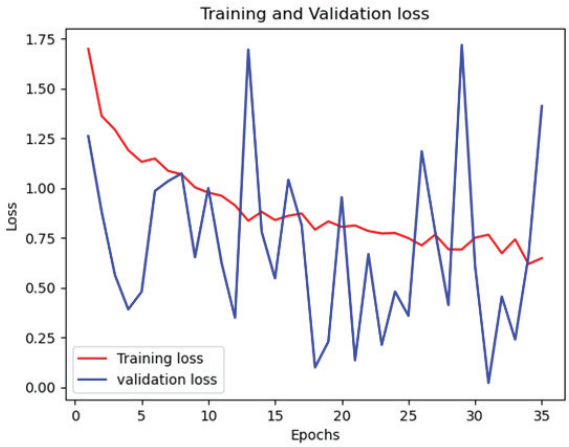

Fig. 15. Training and Validation Loss

Furthermore, the effectiveness of this model can be increased further by training the test dataset with another pre-trained model such as Xception, which is 88 MB in size and top-5 accuracy higher up to 0.945 and also model named MobileNetV2 which has similar accuracy to VGG16 but smaller in size. The main reason for selecting VGG16 is due to its higher number of parameters and the stability of the model.

Finally evaluating the results of the survey conducted among 250 random individuals, it depicts that the general public would prefer to have an online platform to ensure the safeguard of beloved pet dogs.

The main findings of this research project are targeting the emotional aspects of the dog and the owner. There are a lot of abandoned dogs thriving to find a new home or a shelter to themselves. There are a large number of owners who have misplaced their dogs looking for them using various methods. The main purpose of this application is to make a bridge between the two parties.

V. Discussion

The results of this paper highlight only some of the specific objectives achieved. This can be easily used with various other parameters. The results obtained from testing Class 0-4 gave an accuracy of more than 90% by using CNN. The reason behind this could be using a more sophisticated VGG16 model and using of sufficient amount of data for the training. CNNs work well with a large number of data. It was convinced that the accuracy increases in the model with the number of epochs. At a certain point, this was tested with a human face image, to found out that it was given a result that is not confined to only one particular class. That means the dogs from different classes are shown with a higher distance. Later it was rectified by handling the error properly to display an error message.

VI. Conclusion

This research is about image classification by using a combined approach of Deep Learning and Machine Learning through the TensorFlow framework. The objective of the research was to identify or recognize the lost dogs with the usage of image processing technology, which was successfully achieved through this research. CNN becomes the main agenda for this research. The role of the number of epochs is to increase the accuracy of the model and avoid the risk of overfitting. Finally, as mentioned, 5 classes are trained at its most and to enhance, improve and use in solving the real-world problem, all most all the dog breeds or dogs' types should be classified accordingly and feed into the model, which is practically a huge task.

References

[1] R. SAGAR, "What Is The Difference Between Computer Vision And Image Processing?," 26 12 2018. [Online]. Available: https://analyticsindiamag.com/what-is-the-difference-between-computer-vision-and-image-processing/.

[2] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based Learning Applied to Document Recognition," Proc. of IEEE, 1998.

[3] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012.

[4] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende and D. Wierstra, "A Recurrent Neural Network For Image Generation".

[5] M. Rastegari, V. Ordonez, J. Redmon and A. & Farhadi, "XNOR-net: Imagenet classification using binary convolutional neural networks.," 2016.

[6] P. Kamavisdar, S. Saluja and S. & Agrawal, "A survey on image classification approaches and techniques," in International Journal of Advanced Research in Computer and Communication Engineering, 2013.

[7] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici and E. W. J., " SVM active learning approach for image classification using spatial information," IEEE Transactions on Geoscience and Remote Sensing, 2014.

[8] "Flickr|Explore," Flickr, [Online]. Available: https://www.flickr.com/explore.

[9] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," 2018 International Interdisciplinary PhD Workshop (IIPhDW), 2018.