Cat Breeds Classification using Convolutional Neural Network for Multi-Object Image

Naura Qatrunnada [a,1], M. Fachrurrozi [b,2*], Alvi Syahrini Utami [b,3]

[a] Undergraduate Student of Informatics Engineering, Universitas Sriwijaya, Indonesia
[b] Computer Science Faculty, Universitas Sriwijaya, Indonesia
[1] 09021181722067@students.unsri.ac.id; [2*] mfachrz@unsri.ac.id; [3] alvi_syahrini@ilkom.unsri.ac.id
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Image Classification
Multi-Object Image
Convolutional Neural Network

ABSTRACT

Cat is one of the most popular pets. There are many cat breeds with unique characteristic and treatment for each breed. A cat owner can have more than one cat, either the same breed or different breeds. But not all cat owners know the breeds of their cats. Computers can be trained to recognized cat breeds, but there are many challenges for computers because it limited by how much they have been trained and programmed. In recent years, a lot of research about image classification has been done before and got various result, but most of the data used in previous research were single object images. Therefore, this study of cat breeds classification would be conducted with Convolutional Neural Network (CNN) in the Multi-Object images. This method was chosen because it had good classification results in the previous studies. This study used 5 breeds of cats with every breed having 200-3200 images for training. The test results were measured using confusion matrix, obtaining the precision, recall, f1 score and accuracy of 100% on multi-object images with 2 objects and 3 objects. On images with 4 objects achieved the precision, recall, f1 score and accuracy value of 89%, 87%, 87% and 95%. While the value of precision, recall, f1 score and accuracy on images with 5 objects get 87%, 86%, 86% and 94%, respectively.

1. Introduction

Pets can not only help to relieve stress but also as a form of social support that can improve health [1]. One of the most popular pets is the cat. According to the Fédération Internationale Féline (FIFe), there are 48 cat breeds in the world. Each cat has unique characteristics and different health conditions, so it is necessary to provide proper care and treatment. A cat owner can have more than one cat, either the same breed or different breeds. But not all cat owners know the breeds of their cats. Therefore, this study will be discussed the classification of cat breeds.

Research related to the animal breeds classification has been done previously, the study of image classification based on the shape of the cat's face using the Haarcascade and Viola Janes method [2], Recognition of the Scottish Fold using the Histogram of Oriented Gradients method and artificial neural networks [3]. The classification of dog breeds using the Convolutional Neural Network (CNN) method also has been done. The purpose of this research was to recognize different types of dog breeds and obtained an accuracy of 96.75% [4]. However, all of these studies only used single object image data.

Image recognition can not only recognize an object but also can classify multiple objects in an image at once. The study of Content-Based Image Retrieval for Multi-Objects Fruits Recognition using k-Means and k-Nearest Neighbor [5] discussed the CBIR method for recognize fruits and used the k-NN method for classification. This study gained results of 92.5% accuracy on single object images and 90% on multiple object images.

The study of A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning [6] built a mobile application that can recognize cat breeds. The method used is Convolutional Neural Network (CNN) with MobileNet architecture. This application can recognize cat breeds with an average classification accuracy for single object images is 84.74% and 60% for multiple object image detection. The study of Multi-Object Classification and Unsupervised Scene Understanding Using Deep Learning Features and Latent Tree Probabilistic Models [7] combined the Tree Model and Deep Learning methods. With the method they proposed, there is a significant increase in precision, recall and f1 score compared to the artificial neural network classifier consisting of 3 layers.

Convolutional Neural Network (CNN) is a method that applies an artificial neural network algorithm with a deeper layer to process multiple-dimensional data [8]. When training a model of Convolutional Neural Network (CNN) requires a large number of datasets to improve the accuracy and performance of the model. Not only the quantity of the data but also the quality of the data is very important for Convolutional Neural Network (CNN) model performance [6]. Based on the description above, in this study, a classification of multiple object images will be done using the Convolutional Neural Network method.

2. Literature Study

a. Related Works

Multi-Object image classification was also done in the study of Content-Based Image Retrieval for Multi-Objects Fruits Recognition using k-Means and k-Nearest Neighbor [5]. This study discussed the CBIR method for recognizing fruits and the k-NN method for classification. This study gained results of 92.5% accuracy on single object images and 90% on multiple object images.

The study of A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning discussed the detection of cat breeds using Deep Learning [6]. The model used in this study was Mobilenet_v1 FPN and achieved an accuracy of around 81.74% with single object image and 60% for multiple object detection. This research focuses on creating user-friendly applications, so the researcher prioritizes the processing time.

Multi-Object image recognition was also done in the study of BVCNN: A Multi-Object Image Recognition Method Based On The Convolutional Neural Networks [9]. This study proposed the combination of BING and Vectorization of Convolutional Neural Network method. The BING method is a segmentation method and was chosen because it is the fastest speed calculation and small orders of magnitude for target candidate windows that are used to image recognition. Meanwhile, the used of vectorization was to improve the speed of the depth of the Convolutional Neural Network. This proposed method gained the average time required to recognize an image is less than one second. However, the multiple object classification did not achieved the desired results, because there was an error when the model tried to recognize the objects from target candidate windows.

b. Multi-Object Image

In everyday life, humans are able to see and identify many objects in a landscape. According to the Oxford Dictionary, object means a thing that can be seen but is not alive, etc., while multi-means more than one or many. So, multi-object image is an image that contains more than one or many objects that can be recognized.

Objects are treated as part of the image that passes through a classifier for identification. The problem with multiple object images lies in the representation of the image in such a way that integrates information about the image so that there are similar characteristics between one image and another [10]. An image can also have different object positions, rotation, translation, and dilatation [11].

Each image has distinctive characteristics of different colors, shapes and textures. This characteristic can be extracted and used by the system for training so that one image can be distinguished by the other image to be processed [5].

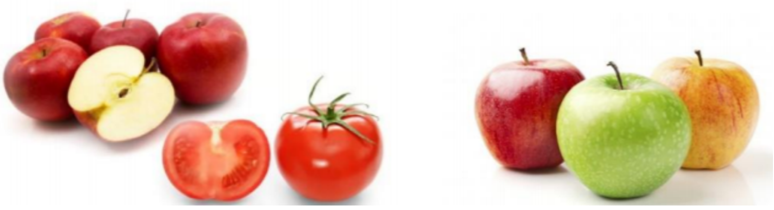

Fig. 1. An example of Multi-Object image

Fig. 1. has more than one different object. The image that shows apples and tomatoes have similar colors and shapes. To distinguish red apples from red tomatoes requires an approach that can identify objects of the same color and shape by utilizing differences in texture. While the image showing red apples, green apples, and orange apples requires an approach that can detect Region of Objects on different apple objects even though they have similar shapes [12].

c. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the algorithms of Deep Learning based on artificial neural network that is used to detect and recognize objects in the digital images [13]. Convolutional Neural Network (CNN) is also the development of a Multi Layer Perceptron that designed to process two-dimensional data [14].

Convolutional Neural Network (CNN) implemented a mathematical operation called convolution on its networks. Convolution is a linear operation that replaced matrix multiplication at least one layer in the network [15]. Generally, Convolutional Neural Network (CNN) consists of convolution layer, pooling layer and fully connected layer.

1) Convolution Layer

The convolution layer is the most important component in Convolutional Neural Network (CNN) consisting of a set of filters. Filter is a matrix usually 2 x 2 or 3 x 3 that convoluted with input to get features on the image and generate feature maps [8].

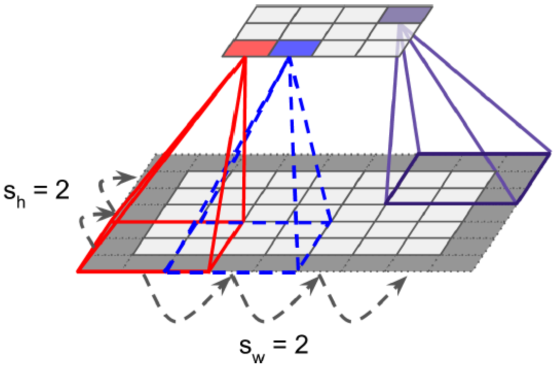

Fig. 2. An example of convolution operation

Fig. 2. is an example of convolution operation given a 5 x 7 input layer with padding to be 0 and convoluted with a 3 x 3 filter with a stride of 2. Stride is the parameter of the number of filter shifts. Calculations are performed at each step when the filters are transferred to activation map input and multiplied by corresponding areas in the activation map to calculate values on the activation map output. The resulting values will be summed up to get the corresponding results (shown in red and blue colors) in the activation map output on each convolution step.

Convolution operation can be written with the following calculation formula:

$$h' = \lfloor \frac{h-f+s}{s} \rfloor , w' = \lfloor \frac{w-f+s}{s} \rfloor \qquad (1)$$

where,

h’, w’ = output value
h, w = height and weight value
f = filter
s = stride

2) Pooling Layer

The purpose of the pooling layer is to increase spatial invariance by reducing the size of the feature map [16]. The pooling layer also aims to reduce the computational resources required to process data and reduce the risk of overfitting [17]. There are two types of pooling operations, max pooling and average pooling. In max pooling operation, calculations are done by getting the maximum value. While the average pooling takes the average value on each row and columns on the feature map.

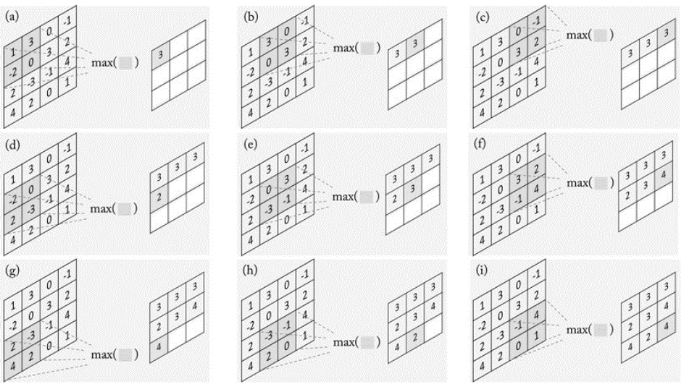

Fig. 3. Max pooling operation

3) Fully Connected Layer

The feature map generated by the previous stage is still in the form of a multi-dimensional array. Therefore, it needs to be transformed first into a one-dimensional array before entering the fully connected layer [14]. This layer transforms the feature map generated by the previous layer in the form of a multiple-dimensional array into a one-dimensional array [17]. The fully connected layer performs the same operation as the convolution layer with a 1 x 1 filter. The difference between the convolution layer and the fully connected layer is that each unit in the convolution layer is only connected to a certain region of the input, whereas each unit in the fully connected layer is connected. to all units from the previous layer [8]. The operation of the fully connected layer can be represented as a simple matrix product as follow:

$$y = f(W^T x + b) \qquad (2)$$

Where x and y are the input and output activation vector values, while W is the matrix containing the weights between the connections of each layer unit, and b is the bias vector value..

d. Convolutional Neural Network Architectures

Convolutional Neural Network (CNN) architecture includes network depth, the dimensions of each layer and the layout of layers can be used as a pre-trained model. Convolutional Neural Network (CNN) has many architectures that can be used, such as ResNet, Xception, VGG16, etc. Xception is one of the CNN architectures which consists of 36 convolutional layers which are divided into 3 main parts: the entry flow, the middle flow (repeated 8 times) and the exit flow. Each network module is separated by a residual network that connects each module as shown in Fig. 4.

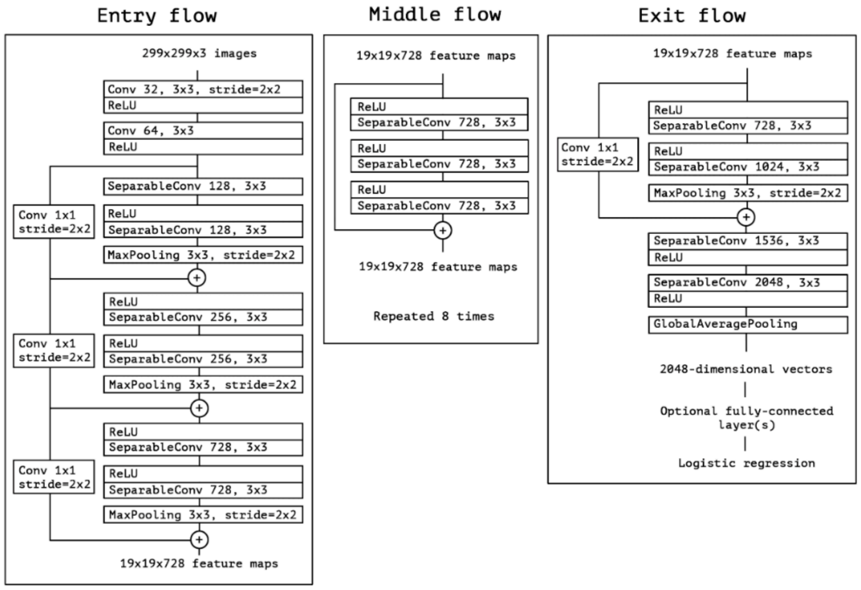

Fig. 4. Xception network architecture

After the first convolutional layer is done, the input will be entered into a separable convolution network consisting of depthwise convolution and pointwise convolution. Each channel of input will be separated spatially by performing a depthwise convolution operation. Then the output is captured by a pointwise convolution that runs a 1x1 convolution and continues on processing to the next convolutional network [18].

3. Methodology

a. Data Collection

The data is obtained from https://www.kaggle.com/ma7555/cat-breeds-dataset by accessing it online from sites that provide image datasets. The Cat Breeds Dataset contains a collection of cat images captured using the PetFinder API. The data is a type of secondary data obtained by downloading the dataset that has been provided by the dataset provider. The downloaded dataset is still in rar format which then extracted to obtain an image in jpg format that has been labeled with the appropriate cat breed. This study used 5 type of cat breeds: Bengal, Calico, Persian, Siamese and Sphynx. Those type of cat breeds were chosen because it has a difference in color and pattern between each other. These breeds also gained the highest accuracy in the previous studies [3, 6, 19]. The data used for training is single object images. The sample image and distribution of data can be seen in Fig. 5. and Table 1.

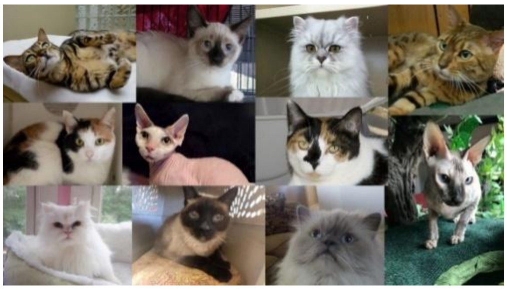

Fig. 5. Sample dataset

Table 1. Cat Breeds Distribution

Label	***Contains***
Bengal	2477
Calico	3468
Persian	4018
Siamese	2888
Sphynx	209
Total	**13060**

b. Reasearch Framework

In this study, the framework is divided into 3 stages. There are pre-processing, training and testing. Before entering the training stage, the training data will be processed first. Firstly, the dataset will be divided into three separate folders specifically are 80% as training data, 10% as validation data and 10% will be used as test data. The pre-processing stage aims to equalize the size and enrich the data. This stage includes the process of resize and augmenting data such as mirroring, rotation and shifting. Because of the limited data, data augmentation is essential to improve the accuracy of the model because it can provide more data for the model to learn so the model become more robust.

The process is carried out to build a Convolutional Neural Network model that can meet research needs. This stage is carried out by entering training data held to find weight optimization and error reduction in the model with the aim of being able to classify cat breeds. After getting the model from the training process, the model will be used for testing. In the testing process is done by entering test data. Then the system will find and recognize the cat's face and compare it with the model to get the result. The process of training and testing can be seen in the following figure.

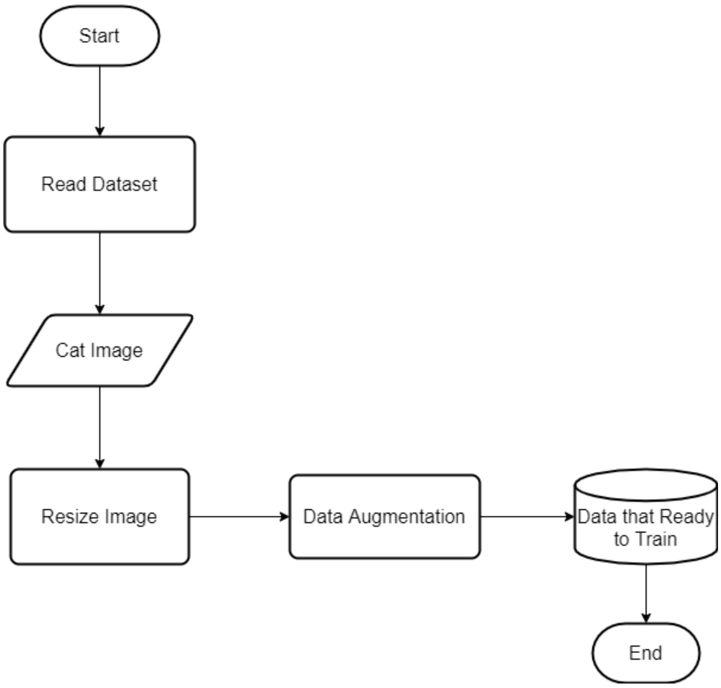

Fig. 6. Pre-Proccesing Process

The training process aims to build a Convolutional Neural Network (CNN) model that can meet the research needs. This training stage was done by entering the training data which is then processed to find weight optimization and decrease the error value in the model with the aim of being able to classify cat breeds. This training process takes a considerable time, depending on the amount of data being trained and the specifications of the computer used.

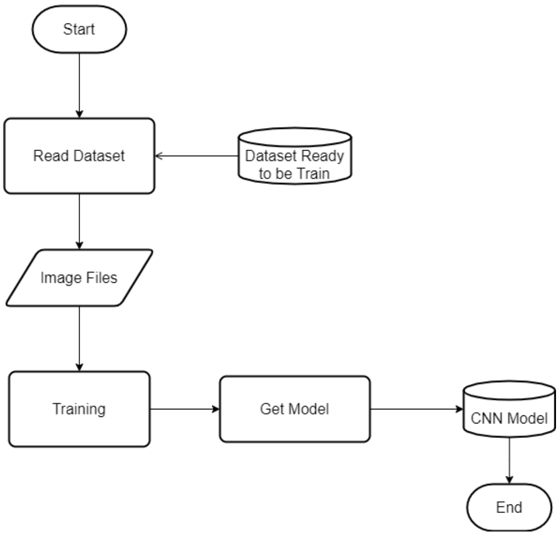

Fig. 7. Training Process

After obtaining the model from the training process, the model will be used for testing. The model that has been obtained from the training process will be used in the testing stage. To help recognize the cat's face, the OpenCV library was used. Then the system will search and detect the cat's face and compare it with the model to get results in the form of cat breeds class. In testing for multiple object images, the steps used are the same. However, the system will store each detected cat face into an array and then compare the cat's face with the model to obtain classification results. The process at the testing stage can be seen in Fig. 8.

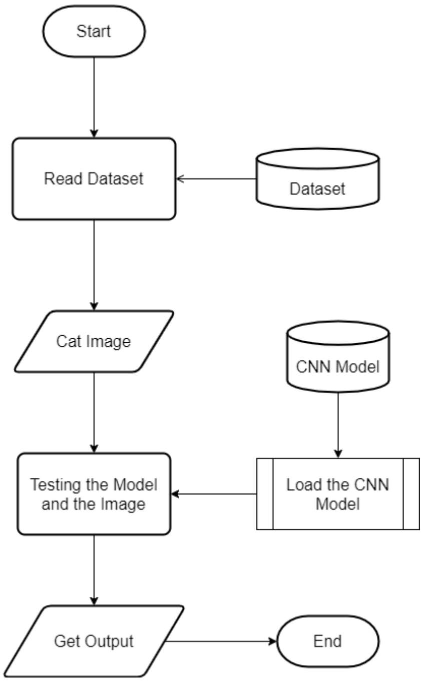

Fig. 7. Testing process

4. Result

The testing of the training model for the cat classification process is divided into two scenarios. The first scenario is to classify single object images. Then in the second scenario, the testing process is carried out on a multi-object image. The test results from the test scenario will be presented using a confusion matrix.

The first scenario involves the image of a cat with only one object. The data consists of five classes with a total of numbers 1312 testing images. Test results from the cat breeds classification process on the single objects images can be seen in Table 2.

Table 2. Single Object Image Confusion Matrix

Class	***Precision***	***Recall***	***F-measures***	***Accuracy***
Bengal	0,90	0,91	0,91	0,97
Calico	0,93	0,90	0,92	0,96
Persian	0,96	0,98	0,97	0,98
Siamese	0,92	0,94	0,93	0,97
Sphynx	0,95	0,91	0,93	1,00

As can be seen in the Table 2, all of the five cat breeds got the precision, recall, f1 score and accuracy of above 90%. Each of these cat breeds has its own unique characteristic. For example, a Persian cat has a snub nose and long fur and the Sphynx is a hairless cat. The results revealed that this model is effective at extracting cat breeds with unique features.

The second scenario test is performed on multiple object images. The images used are 140 images of cats that were randomly selected from the test data in the first scenario and the internet which were combined in such a way to produce 40 multiple object images. The tested images consist of 10 multiple object images each with 2 objects, 3 objects, 4 objects and 5 objects (Fig. 9.) The results of the second scenario test are also displayed in the confusion matrix calculation format which is presented in Table 3.

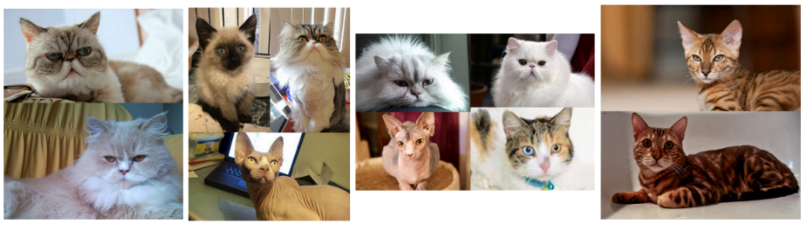

Fig. 9. Sample of multi-object image

Table 3. Multi-Object Image Confusion Matrix

Category	*Precision*	*Recall*	*F1 Score*	*Accuracy*
2 Objects	1,00	1,00	1,00	1,00
3 Objects	1,00	1,00	1,00	1,00
4 Objects	0,89	0,87	0,87	0,95
5 Objects	0,87	0,86	0,86	0,94

The second test scenario is carried out on multi-object images that have more than one different or similar cat breed. The system will recognize the cat's face as an object and classify the object. Then display the results of the cat race class that the object belongs to. Tests on images consisting of 2 objects and 3 objects get the same overall value of precision, recall, f1 score and accuracy, which is 100%. The system can recognize and classify cat breeds correctly. Tests on images consisting of 4 objects get precision, recall, f1 scores and accuracy values of 89%, 87%, 87% and 95%, respectively. While the image that has 5 objects gets precision values, recall, f1 score and accuracy of 87%, 86%, 86% and 94%. If there are more than 3 objects in an image, it makes the quality of the image size to decrease so that the system will be mistaken in recognizing cat breeds.

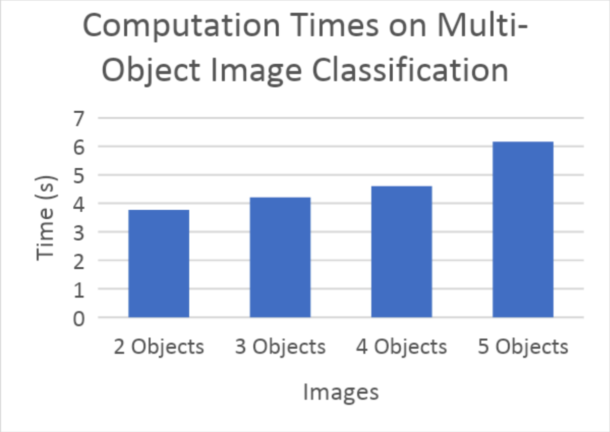

Fig. 10. Sample of multi-object image

In addition, the computational time in classifying began to decrease. As can be seen in Table 3 and Fig. 10., the accuracy and computation time will increase if the image has fewer objects. The average computation time required for 2 objects is 3.76 second. While the average computation time required for 3 objects and 4 objects is 4.20 seconds and 4.60 seconds, respectively. An image with 5 objects requires a longer computation time, which is 6.16 seconds. This is because the number of objects affects the system when it recognizes a cat's face. In addition, uncertain conditions such as lighting, position, color and background also affect the system in classifying cat breeds.

5. Conclusion

The conclusions obtained based on the results of the system developed regarding the classification of cat breeds with Convolutional Neural Network on multiple object images are as follows: Classification of cat breeds on multiple object images can be done using the Convolutional Neural Network method with Xception Architecture. The implementation of the Convolutional Neural Netwok method on multi-object images produce a precision values, recall¸ f1 score and an accuracy of 100% on images with 2 objects and 3 objects. In images with 4 objects get precision values, recall, f1 score and accuracy of 89%, 87%, 87% and 95% respectively. While the precision value, recall, f1 score and accuracy on the image with 5 objects get 87%, 86%, 86% and 94%. respectively. The more objects in an image, the value of accuracy will start to decrease. Vice versa, the fewer objects, the accuracy value will increase. Computing time is also affected by the number of objects contained in the image. The more objects there are, the longer of the computation time.

References

[1] W. C. Compton and E. L. Hoffman, "Positive Psychology: The Science of Happiness and Flourishing," in SAGE Publication, 2019.

[2] M. R. Effendi, "*Sistem Deteksi Wajah Jenis Kucing dengan Image Classification menggunakan OpenCV*," 2016

[3] S. Indriyani, F. Sthevanie & K. N. Ramadhani, "*Pengenalan Ras Kucing Scottish Fold Menggunakan Metode Histogram of Oriented Gradients dan Jaringan Saraf Tiruan*," 6(2), 9325 – 9335, 2019.

[4] Borwarnginn, P., Thongkanchorn, K., Kanchanapreechakorn, S., & Kusakunniran, W., "Breakthrough Conventional Based Approach for Dog Breed Classification Using CNN with Transfer Learning." in 11th International Conference on Information Technology and Electrical Engineering, ICITEE 2019, 7, 1-5, 2019. https://doi.org/10.1109/ICITEED.2019.8929955

[5] M. Fachrurrozi, A. Fiqih, B. R. Saputra & R. Algani, "Content Based Image Retrieval for Multi-Objects Fruits Recognition using k-Means and k-Nearest Neighbor," 1–6, 2017.

[6] X. Zhang, L. Yang & R. Sinnott, "A Mobile Application for Cat Detection and Breed Recognition Based on Deep Learning," in AI4Mobile 2019 - 2019 IEEE 1st International Workshop on Artificial Intelligence for Mobile, 7–12, 2019.

[7] T. Nimmagadda & A. Anandkumar, "Multi-Object Classification and Unsupervised Scene Understanding Using Deep Learning Features and Latent Tree Probabilistic Models", 2015.

[8] S. Khan, H. Rahmani, S. A. A. Shah & M. Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision," in Synthesis Lectures on Computer Vision, 8(1), 2018.

[9] H. Shi & S. Wang, "BVCNN : a multi-object image recognition method based on the convolutional neural networks", https://doi.org/10.1109/ICVRV.2015.28, 2015.

[10] M. Dimitriou, T. Kounalakis, N. Vidakis & G. Triantafyllidis, "Detection and Classification of Multiple Objects using an RGB-D Sensor and Linear Spatial Pyramid Matching," 12(2), 78–87, 2013.

[11] P. Hao, Y. Ding & Y. Fang, "Image Retrieval Based On Fuzzy Kernel Clustering and Invariant Moments", https://doi.org/10.1109/IITA.2008.189, 2008.

[12] R. Khan & R. Debnath, "Multi Class Fruit Classification Using Efficient Object Detection and Recognition Techniques", https://doi.org/10.5815/ijigsp.2019.08.01, 2019.

[13] I. F. Alam, M. I. Sarita & A. M. Sajiah, "Metode Convolutional Neural Network," 5(2), 237–244, 2019.

[14] W. S. Eka Putra, "Klasifikasi Image Menggunakan Convolutional Neural Network (CNN) pada Caltech 101," in Jurnal Teknik ITS, 5(1), 2016.

[15] I. Goodfellow, Y. Bengio & A. Courville, "Deep Learning. MIT Press", 2016.

[16] D. Scherer, M. Andreas & S. Behnke, “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition,” 2010.

[17] A. Géron, “Hands-On Machine Learning with Scikit-Learn (1st Ed),” from O’Reilly Media, 2017.

[18] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", 2014.

[19] M. Afif, A. Fawwaz, K. N. Ramadhani & F. Sthevanie, "Klasifikasi Ras pada Kucing menggunakan Algoritma Convolutional Neural Network (CNN)," in 8(1), 715–730, 2021.