ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

# Computer Vision: Classification of Images Based On Deep Learning with the CNN Architecture Model

[1] Heru Syah Putra, [2] Aji Tri Pamungkas Nurcahyo, [3] Haryanto, [4] Chuang-Jan Chang, [5] Shu-Lin Hwang

[1] [2] [3] [4] [5] Ming Chi University of Technology, Gongzhuan Road, Taishan District, New Taipei City, Taiwan.
Corresponding Author Email: [1] herusyahputra@unib.ac.id, [2] ajipamungkastrinurcahyo@gmail.com, [3] anto112@gmail.com, [4] cjchng@mail.mcut.edu.tw, [5] hwangsl@mcut.edu.tw

***Abstract*— Deep learning is a scientific field in Machine Learning (ML) that is developing with various applications, one of which is visual image processing technology. With the excellent capabilities of computer vision, image processing from computer visuals is used to duplicate the human ability to understand object information in the image. One of the Machine Learning (ML) methods that can be used for object classification in images is the Convolution Neural Network (CNN) method. The two core stages when processing object classification in the image, the first stage is image classification using feedforward, and the second stage applies the backpropagation method. In this study, before the classification stage, this method was first carried out through preprocessing, which is useful as an image separation to focus on the object to be classified. Furthermore, it is carried out by conducting pre-training using the feedforward method with the bias weights, which are updated after every training process. The observations of this study, the results of image classification training with a degree of ambiguity, resulted in a good average accuracy validation value of 0.91 in a confidence interval with a range of 0-1. So, it can be concluded that applying the Convolution Neural Network (CNN) method to distinguish objects in an image can classify them well.***

***Index Term:* Computer Vision, Image Processing, Deep learning, CNN Architecture Model**

## I. INTRODUCTION

In today's digital life, digital images are everywhere around us. An image is a visual representation of an object, person, animal, or scene. A digital image is a two-dimensional function f(x, y), a 3-dimensional view into a 2-dimensional projection plane, where x, y represents the location of an image element or pixel and contains an intensity value. If the values of x, y and intensity are discrete, then the image is said to be a fortified image. A digital image is a matrix representation of a two-dimensional image using a finite number of points of cell elements, commonly referred to as pixels (image elements, or mops).

One of the visual problems in computer vision that has long been sought for a solution is the classification of objects in the image in general, how to duplicate the ability of humans to understand image information so that computers can recognize objects in the image like humans. The feature engineering process used in general is very limited, which can only apply to certain datasets without the ability to generalize any type of image. This is due to differences between images, including differences in angle of view, differences in scale, differences in lighting conditions, deformation of objects, and so on.[1]

Academics who have been struggling with this problem for a long time. One of the successful approaches used is to use Artificial Neural Networks (JST), which are inspired by neural networks in humans. The concept was then further developed in Deep Learning. In 1989, Yann LeCun and his friends successfully performed zip code image classification using a special case from Feed Forward Neural Network under Convolution Neural Network (CNN). Due to hardware limitations, Deep Learning was not further developed until 2009. Jurgen developed a Recurrent Neural Network (RNN) that gained significant results in handwriting recognition. Since then, with the development of computing in the hardware of the Graphical Processing Unit. (GPU), DNN development is running at a rapid pace. In 2012, a CNN could perform image recognition with an accuracy that rivals that of humans on a given dataset. Today, Deep Learning has become one of the hot topics in the world of Machine Learning because of its significant ability to model complex data such as imagery and sound. [2][3][4]

The Deep Learning method that currently has the most significant results in image recognition is the Convolutional Neural Network (CNN). This is because CNN is trying to imitate the image recognition system in the visual cortex [5][6]humans to process image information. But CNN, like other Deep Learning methods, has the disadvantage of a long model training process. With the development of hardware, this can be overcome using General Purpose Graphical Processing Unit (GPGPU) technology.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

## II. RELATED WORK

### A. Convolutional Neural Network

Convolutional Neural Network (CNN) is a Multilayer Perceptron (MLP) development designed to process two-dimensional data. CNN belongs to the Deep Neural Network type due to the high network depth and is widely applied to image data. In the case of image classification, MLP is less suitable for use because it does not store spatial information from image data and considers each pixel to be an independent feature resulting in poor results.

CNN was first developed under the name NeoCognitron by Kunihiko Fukushima, a researcher from NHK Broadcasting Science Research Laboratories, Kinuta, Setagaya, Tokyo, Japan. The concept was later matured by Yann LeChun, a researcher from AT&T Bell Laboratories in Holmdel, New Jersey, USA. LeCchun successfully applied the CNN model under the name LeNet to his research on recognizing numbers and handwriting. In 2012, Alex Krizhevsky, with his CNN implementation, won the ImageNet Large Scale Visual Recognition Challenge 2012 competition. This achievement is a moment of proof that deep learning methods, especially CNN. The CNN method has proven successful in outperforming other Machine Learning methods such as SVM in the case of object classification in images.[5][2].

### B. CNN Concept

The way CNN works has similarities to MLP. However, in CNN, each neuron is presented in a two-dimensional form, unlike MLP, where each neuron is only one-dimensional in size.[7]

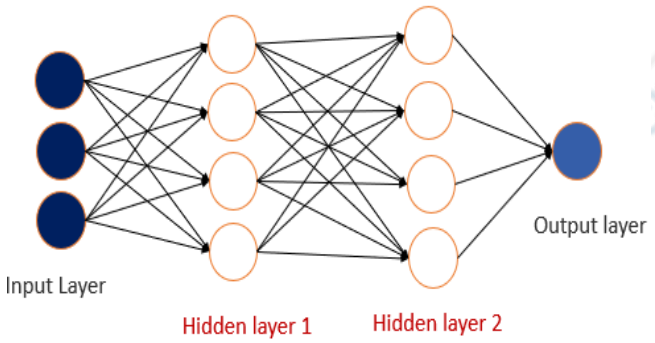


**Figure 1.** Architecture Basic MLP

An MLP, as shown in figure 1. It has an I layer, with each layer containing $_{Ji}$ neurons. MLP accepts one-dimensional data inputs and propagates that data on the network until it produces an output. Each connection between neurons on two contiguous layers has a one-dimensional weight parameter that determines the quality of the mode. In each input data on the layer, a linear operation is carried out with the existing weight value. Then the computational results will be transformed using a non-linear operation called an activation function.

On CNN, the data propagated on the network is two-dimensional, so the linear operation and weighting parameters on CNN are different. On CNN, linear operations use convolution operations, while weights are no longer one-dimensional only. However, they are four-dimensional, which is a collection of convolution kernels, as shown in figure 2. Due to the nature of the convolution process, then CNN can only be used on data that has a two-dimensional structure, such as imagery and sound.

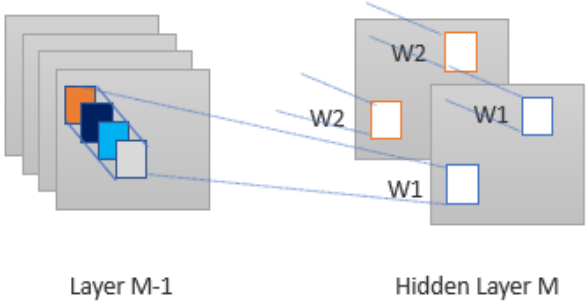


**Figure 2.** Process convolution CNN.

## III. MATERIAL METHOD

We introduced the Ensemble Convolution Neural Network Architecture (CNN), consisting of CNN-based object classification networks and object detection to train datasets taken from two types of animals.

Experimental analysis showed better results than Places-CNNs.

The convolution layer volution pooling acts as a feature extractor of the input image, while the layer that is fully layered acts as a classification.

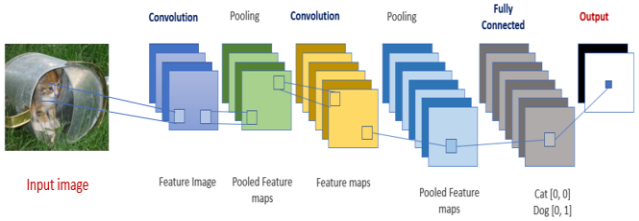


**Figure 3**. Convolution Layer

In the image above, when receiving the desired image as input, the network correctly gives the highest probability for it (0.94) among the four categories. The sum of all probabilities in the output layer should be one. There are four main operations in ConvNet shown Convolution, Non-Linearity (ReLu), Pooling or Sub Sampling, and Classification (Fully Connected Layer[8] ).

### Convolutional Neural Networks – Architecture

The layer is used to build Convolutional Neural Networks. Simple ConvNet is a sequence of layers, and each ConvNet layer converts one activation volume to another volume through the yang function can be distinguished. We used four main layer types to build the ConvNet architect.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

**Convolution Layer**

It holds the raw pixel value of the training image as input. In the above example, the image (cat) with a width of 32, a height of 32, and three color channels, R, G, and B, are used. It passes the forward propagation step and finds the probability of output for the class setup. Lapsang ensures spatial connections between pixels by interpreting image features using a small box of input data. Let's assume the output probability for the image above is [0.2, 0.1, 0.3, 0.4]. The size of the feature map is controlled by three parameters.

- Depth – The number of filters used for coevolutionary operations
- Stride – The number of filters used to filter the matrix above the input matrix.
- Padding – very good for entering matrices with zeros around the limit matrix.

Calculates the total error on the output layer with the summation of all 4 classes.

***Total Error* = ∑ *1/2 (Target probability – Output probability)*$^2$**

Calculated the output of neurons that are connected to the local area inputted. It can produce volumes such as *[32x32x16] for 16 filters.*

**Rectified Linear Unit (ReLu) Layer**

A non-linear operation. This layer implements the function of element-wise activation. ReLu is used after every convolution operation. It is applied per pixel and replaces all negative pixel values in the feature map with zeros. This makes the measure n volume unchanged ([32x32x16]), where ReLu is a non-linear operation.

**Pooling Layer**

As subsampling or downsampling, the pooling layer performs downsampling operations along the spatial dimensions (Width, height), resulting in volumes such as [16X16X16], i.e., reducing the dimensions of each feature map but retaining the most important information.

**Max pooling operation on a Rectified Feature map:**

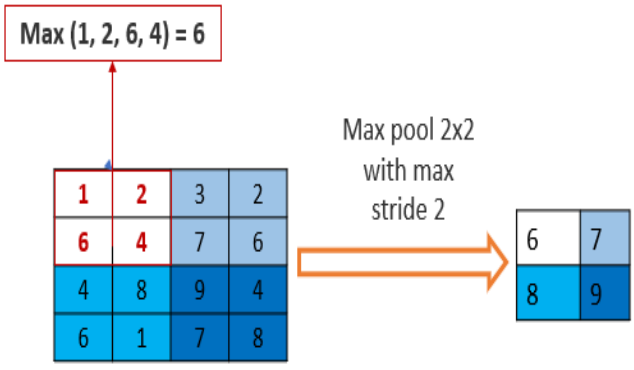


**Figure 4.** Max pooling operation

**Fully Connected Layer**

Inside the Fully Connected Layer, each node is connected to every other node in an adjacent layer. The FC layer calculates the class score with a traditional multilayer perceptron that uses softmax activation conditions in the output layer. This results in a volume size of [1x1x10], in which each of the 10 numbers corresponds to the class score.

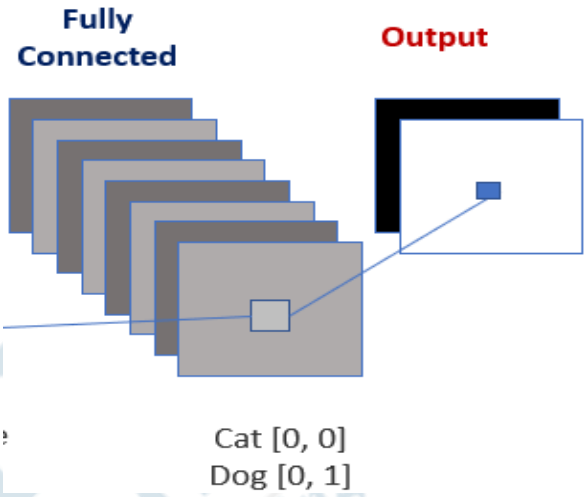


**Figure 5.** Process Fully Connected Layer

## IV. RESULT AND DISCUSSION

In this paper, the dataset image used is cats and dogs, where the data will be used as a sample for recognition and classification in the form of image data.

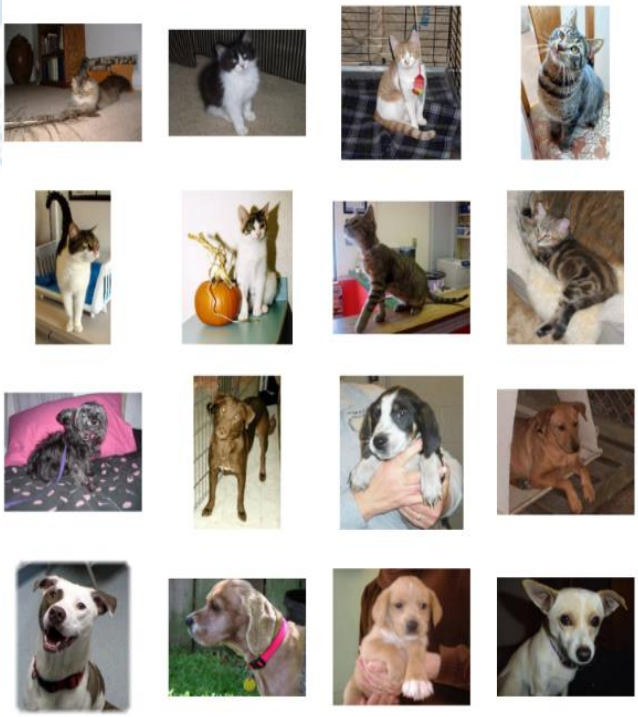

**Figure 6.** Cat and dogs images dataset

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 11, November 2022

### Evaluation of Model

In this stage, the CNN Convolution Neural Network method processes the model dataset images. Training and validation sets are used while training the dataset in the convolutional neural network model, as shown in figure 7 below.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 150, 150, 3)]     0

 conv2d (Conv2D)             (None, 148, 148, 16)      448

 max_pooling2d (MaxPooling2D  (None, 74, 74, 16)       0
 )

 conv2d_1 (Conv2D)           (None, 72, 72, 32)        4640

 max_pooling2d_1 (MaxPooling  (None, 36, 36, 32)       0
 2D)

 conv2d_2 (Conv2D)           (None, 34, 34, 64)        18496

 max_pooling2d_2 (MaxPooling  (None, 17, 17, 64)       0
 2D)

 conv2d_3 (Conv2D)           (None, 15, 15, 128)       73856

 max_pooling2d_3 (MaxPooling  (None, 7, 7, 128)        0
 2D)

 flatten (Flatten)           (None, 6272)              0

 dense (Dense)               (None, 512)               3211776

 dense_1 (Dense)             (None, 1)                 513

=================================================================
Total params: 3,309,729
Trainable params: 3,309,729
Non-trainable params: 0
_________________________________________________________________
```

**Figure 7.** Summarize the model architecture

After that, the visualization process of dataset images on the Neural Network of each layer will be carried out as shown in figure 8 below,

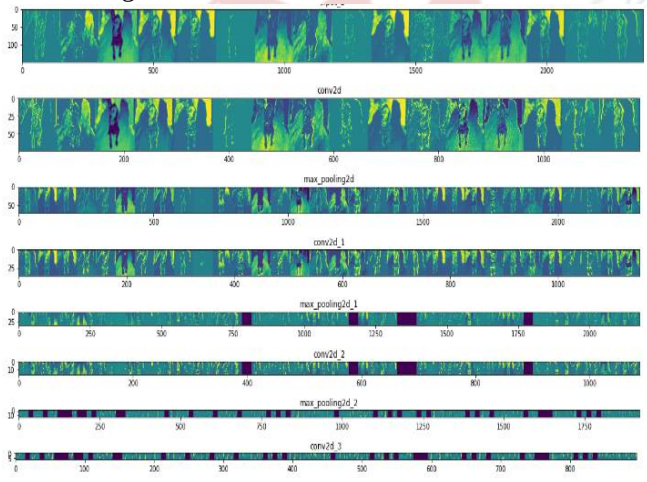


**Figure 8.** Visualization layer

During the classifier process with the Convolution Neural Network CNN model, the evaluation device is used to evaluate the performance of the trained model. Accuracy is measured by the following formula.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

*Where:* TP = True Positive, TN = True Negative, FP = False positive, FN = false negative

### Obtain Result

Results obtained by evaluating a different trained model with a given set of evaluations

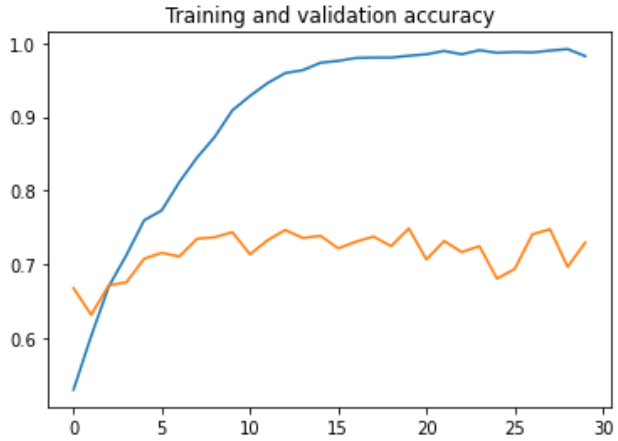


**Figure 9.** Training and Validation Accuracy Result

The curve above, we can see that the proposed model has better accuracy. Since the beginning, the model has been trained to classify the image of the animal.


**Figure 10**. Traing and Validation Loss

We can also see the loss and accuracy curves of the proposed model in figure x and y. Based on the curve, that gap between validation and accuracy.

## V. CONCLUSION

The classification method using Convolutional Neural Networks is reliable enough to determine the correctness of the classification of object images. This is evident in the accuracy results, as shown in the curve of figure 9. The change in the degree of confusion in the process of classifying images does not affect the accuracy of the results.

Proves that the classification using the CNN method is relatively reliable to the changes in the parameters made. By using good and optimal data, the subset of the training data will also produce a good classification.

## REFERENCE

[1] P. G. Tushar Jajodia, "Image Classification - Cat and Dog Images," International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 2, pp. 570-572, Dec 2019.

[2] L. Y, "Bangla Handwritten Character Recognition Using extended Convolution Neural Network," Journal of Computer and Communications, vol. 9, pp. 1-14, March 2021.

[3] A. N. H. L. Adam Coates, "An Analysis of Single-Layer Networks in Unsupervised Feature Learning," Proceedings of Machine Learning Research, pp. 215-223, 2011.

[4] R. T. O. S. H. P. N. D. F. H. M. K. a. D. H. (. C. Zijie J. Wang, "CNN Explainer: Learning Convolution Neural Network Interactive Visualization," arXiv, pp. 1-11, August 2004.

[5] K. Fukushima, "Neocognitron: A self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," Springer-Verlag, pp. 193-202, 1980.

[6] M. V. A. S. H. Muthukrishnan Ramprasath, "Image Classification using Convolution Neural Networks," Internation Journal of Pure and Applied Mathematics, vol. 119, pp. 1307-1319, 2018.

[7] K. K. M. Shanmukhi. K Lakshmi Durga Madela Mounika, "Convolution Neural Network for Supervised Image Classification," International Journal of Pure and Applied Mathematics, vol. 119, pp. 77-83, 2018.

[8] N. B. Timea Bezdan, "Convolution Neural Network Layers and Architectures," International Scientific Conference on Information Technology and Data Related Research, pp. 445-451, 2019.