Photo Content Classification Using Convolutional Neural Network

Yilin Hou

1Polytechnic Institute, Purdue University, West Lafayette, IN, 47906, U.S.
*Corresponding author’s e-mail: hou90@purdue.edu

Abstract. This research is about photo content classification using the TensorFlow framework and the deep neural network algorithm in deep learning. Machine learning is a multidisciplinary major whose main research object is artificial intelligence, which enables machines to have the learning ability like human beings. As a branch of machine learning, deep learning analyses and interprets texts, images, and sounds by learning the rules and representations of sample data. A convolutional neural network is an approach of a deep neural network, which is mostly applied to analyse the photo content and has good performance in photo content classification. It is also the main point that the study focuses on. The Visual Geometry Group 16 (VGG16) model in the convolutional neural network was used to achieve the research objectives and has excellent performance. The research used data from CelebFaces and Kaggle, including photos about people, food, animals, and architecture. The photo content classification model in this study can classify these four types of images. Moreover, the analysis of the work, summary, as well as the future work were provided at the end.

1. Introduction
Image classification is the main field of image analysis, in which deep neural networks play the most important role in image analysis. It is assigning a label to an input image from a set of fixed categories. It can be used in lots of fields, like medicine, security, education [1]. Other tasks, such as target detection, image segmentation, image generation, and video comprehension, are highly dependent on the ability of feature expression in image classification. The most typical example is Facebook. The accuracy of Facebook's facial recognition technology, 98 percent, depends on the classification ability of the images[2].

The main method to implement image classification technology is based on deep learning. Deep learning is an artificial intelligence function that teaches computers to process data and create patterns for use in decision making in the way that the human brain works[3]. Convolutional neural network (CNN), a deep learning algorithm, has achieved unprecedented success in the field of computer vision. It is one of the main categories of image recognition, image classification, target detection, and face recognition. The development of the convolutional neural network in image classification can directly affect the performance of all computer vision tasks based on deep learning. Due to its huge commercial value, deep learning and convolutional neural network have become the research hotspot. At present, many traditional computer vision algorithms have been replaced by deep learning[4].

In Holger's paper, Anatomy-Specific Classification of Medical Image Using Deep Convolutional Nets, an image classification model for human anatomy was established by using Convolutional neural network in Deep learning, which laid a foundation for the establishment of the computer-aided diagnostic system. Deep learning performs better than other methods in image classification tasks.
Holger proposed a method for the specific classification of organs or bodies of medical images using computer tomography (CT) of ConvNets. The model was trained using patient images extracted from the hospital Picture Archiving and Communication System (PACS) archives and combined with data augmentation methods to improve classification performance [5].

The paper about kernel based learning approach for satellite image classification using support vector machine published in 2011 discussed that the model based on Support Vector Machine (SVM) algorithm in machine learning, a supervised machine learning method for classification, is applied to the classification of multispectral satellite image data set, and the low-dimensional feature space is transformed into high-dimensional feature space to find the maximum boundary of the class. The SVM is a powerful kernel-based classifier for solving classification problems. Manthira used LISS-3 and AWIFS sensor data from the Indian remote sensing platform Resourcesat-1. The accuracy of the model is compared with that of the traditional classification methods, and the results show that the satellite image classification based on the support vector machine has a better performance[6].

In this paper, based on the environment of Keras and Tensorflow, Python will be used as the programming language to build a photo content classification model of people, food, animals and architecture. The classification of those four aspects is significant. It is a great way to help users categorize photos on their phones, computers, and cloud disk. Now more and more people choose to go out to travel, visit, and participate in study tours. A large number of photos are taken and saved on their equipment during this process, and it is difficult to find a specific photo among all images later. Image classification can categorize things with uniform features into one category, which will save users much time.

2. Method

The image classification mainly consists of three steps, which are collecting and training input images, implementing the convolutional neural network, and the category that images belong to will be classified in the end. In this research, the VGG16 model was used for image classification. VGG16 is a convolutional neural network model, which contains 13 convolutional layers and ReLU layers, five max-pooling layers, three fully connected layers, and one softmax layer[7]. The architecture of VGG16 was shown in Figure 1. Each kind of layer will be discussed in the following. Before those, the TensorFlow environment should be installed based on Python language.

![Figure 1. The architecture of VGG-16.](image-url)
2.1. Data Preparation
The input data contains thousands of images. These images belong to four categories: people, animals, food, architecture. The images of people are from CelebFaces dataset, and the other three are downloaded from Kaggle. The image dataset name, image type, and the number of images in each category are shown in Table 1.

Table 1. Four Image Datasets

| Number | Type of Images | Data Source          | Number of Images |
|--------|----------------|----------------------|-----------------|
| 1      | People         | CelebFaces           | 202599          |
| 2      | Animals        | Kaggle Animals-10    | 15940           |
| 3      | Food           | Kaggle Food-11       | 9866            |
| 4      | Architecture   | Kaggle Architecture Dataset | 4794 |

Since the computer cannot directly manipulate images, videos, text, etc., the first thing to do is to turn the images into an array of pixels, and make sure the images that will be used are of the same size. The image was scaled to find the optimal value of image size, a standard size that is smaller than the actual image resolution, that not only shows. The output size of images here is 224×224 pixels. The example input photos were shown in Figure 2.

2.2. Implement Convolutional Neural Network
A Convolutional neural network consists of many different filters to create a feature map, ReLU layers to increase non-linearity, pooling layers, and a fully connected layer. The input images are passed through those layers and then generates the output [8].

2.2.1. Convolutional layer. The first one is always the convolutional layer. It is the core block of the convolutional network that does most of the computational heavy lifting [9]. The main purpose of the convolutional layer is to extract image features. It overrides the filter at some point in the image, multiplying the value in the filter by the value of the corresponding pixel in the image. The sum is then the value of the target pixel in the output image [10]. Repeat this operation for all locations of the image. The principle of convolution was shown in Figure 3.
2.2.2. **Relu layer.** The RuLU layer is usually used after each convolutional layer. The purpose of this layer is to increase the non-linearity in the network, and the system basically just computes linear operations in the convolutional layer. In basic terms, this layer just removes all the negative values and sets them to 0 [8]. There are some nonlinear functions like sigmoid, tanh can be used but required a large amount of work, while ReLU activation function saves much computation in the whole process, without affecting the accuracy. It also helps to alleviate the problem of vanishing gradients [11].

2.2.3. **Pooling layer.** Pooling layer is also an important concept in convolutional neural networks. One of the important roles of pooling is to preserve the main features while reducing parameters to prevent overfitting. An image contains a lot of information and features, but some of the information is not too useful or repetitive to do image tasks. Such redundant information is removed, and the most important features are extracted. There are many nonlinear pooling functions of different forms, such as maximum pooling, average pooling, etc., among which "max pooling" is the most common. It divides the input image into several blocks and outputs the maximum value for each block. The pooling layer will continuously reduce the space size of data, so the number of parameters and calculation amount will also decrease, which also controls overfitting to a certain extent. The pooling layer is generally placed behind the convolution layer. Therefore, the pooling layer pools the output of the convolution layer [12]. Max pooling layer was shown in Figure 4.
2.2.4. Fully connected layer and Softmax layer. The fully connected layer is the result of flattening the convolutional and pooling process into a single value vector for image classification. Each of its nodes is connected to all of the nodes at the next level. The input value is multiplied by the weight and passed through an activation function (usually ReLU), which is then passed to the output layer. In the output layer, each neuron represents a classification label [13]. After passing through the full connection layer, the last layer uses the Softmax activation function, the result of which can be used to interpret the probability distribution [14].

2.2.5. VGG16 model. The main method or model that was used in this research is the VGG16 model in a convolutional neural network. It contains 13 hidden layers and three fully connected layers. It uses three 3×3 convolution layers instead of a 7×7. The 3×3 filter can maintain the same perceived field as 7×7, but it has more nonlinear transformation with fewer parameters. The convolution of 1×1 is effective, but it is not as good as that of 3×3, which can learn more features. After each convolution layer, a pooling layer is connected to reduce the dimensionality and extract the main features of the image. It simplifies the neural network. However, at the same time, the deep network may cause the problem of vanishing gradient. It also requires a lot of parameters and much time to train.

3. Result and Discussion
The accuracy of the model is 99.67% after training and testing, which means when users upload images to do the classification using this model, the probability that they will get correct classification is more than 99 percent. The training and validation loss/accuracy was shown in Figure 5.

Figure 5. (a) represents the training and validation accuracy. (b) represents the training and validation loss. In both of them, the blue points are training accuracy and the red dotted lines are validation accuracy.
After finishing training the model, several pictures were randomly selected from the training set for classification test, and the classification results were 100% correct. The classification result was shown in Figure 6. Although the photos shown below are squares, the model can classify photos of different sizes or not square shapes. It will reshape the photos for better classification in the process of classifying.

![Classification Results](image)

Figure 6. (a) is the result of animal photos classification, (b)(c)(d) are results of people, food, and architecture classification

When CNN is used to realize image classification, the accuracy rate exceeds 99%, which may be due to the large amount of data and relatively distinct data features. Sharp image features will help improve classification accuracy.

The VGG16 model provides high accuracy. When the layer and distribution changed, the training process took about the same time, but the accuracy was much lower than VGG16. A lower network will cause a decrease in accuracy, but a too deep network will also cause the gradient vanishing due to the unstable full-time network update. If the activation function is changed from ReLU to Sigmoid, the model will also be greatly affected. The training speed of Sigmoid is slow, because the derivative of Sigmoid will be very small. Therefore, the weight is basically no longer updated, which causes the problem of gradient vanishing [15].

4. Conclusion

In summary, this paper proposes an photo content classification model based on TensorFlow framework by using the CNN algorithm in deep learning and completes the research objective of this paper. The model has achieved good results. Deep learning and neural network are the focus of this paper, mainly focusing photo content image classification because of the excellent performance of deep neural networks in that. The model is trained with the input image data, and finally, the images to be classified are classified into one kind of people, animals, food, and architectures. The final accuracy of the model can reach 99 percent, almost accurately categorizing the images into their categories.
In the future, more methods and details will be explored to achieve better performance. For example, more aggressive data augmentation will be helpful. It can help to solve overfitting problem. Overfitting can also be solved by adding dropout layers or trying different architecture such as VGG19 and DenseNet121. Several parameters, such as the number of epoch and the activation function, can also be changed to modify and improve the algorithm.

References
[1] Gavali, P., & Banu, S. J. (2019). Deep Convolutional Neural Network for Image Classification on CUDA Platform. In A. K. Sangaiah (Eds.), Deep Learning and Parallel Computing Environment for Bioengineering Systems. Academic Press, St. Louis. pp. 99-122
[2] What is the Working of Image Recognition and How it is Used? (2020, June 08). Retrieved June 25, 2020, from https://marutitech.com/working-image-recognition/
[3] Hargrave, M. (2020). How Deep Learning Can Help Prevent Financial Fraud. Retrieved June 25, 2020, from https://www.investopedia.com/terms/d/deep-learning.asp
[4] Saha, S. (2018). A Comprehensive Guide to Convolutional Neural Networks-the ELI5 way. Retrieved June 25, 2020, from https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
[5] Roth, H. R., Lee, C. T., Shin, H., Seff, A., Kim, L., Yao, J., . . . Summers, R. M. (2015). Anatomy-specific classification of medical images using deep convolutional nets. https://ieeexplore.ieee.org/abstract/document/7163826
[6] Moorthi, S. M., Misra, I., Kaur, R., Darji, N. P., & Ramakrishnan, R. (2011). Kernel based learning approach for satellite image classification using support vector machine. https://ieeexplore.ieee.org/document/6069282
[7] Muneeb ul Hassan (2018). VGG16 - Convolutional Network for Classification and Detection. https://neurohive.io/en/popular-networks/vgg16/
[8] Bonner, A. (2019, June 01). The Complete Beginner's Guide to Deep Learning: Convolutional Neural Networks. https://towardsdatascience.com/wtf-is-image-classification-8e78a8235acb
[9] CS231n Convolutional Neural Networks for Visual Recognition. (n.d.). https://cs231n.github.io/convolutional-neural-networks/
[10] Sorokina, K. (2019). Image Classification with Convolutional Neural Networks. https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8
[11] Deshpande, A. (2016). A Beginner's Guide To Understanding Convolutional Neural Networks Part 2. https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/
[12] SavyakhoslaCheck out this Author's contributed articles., Savyakhosla, & Check out this Author's contributed articles. (2019). CNN: Introduction to Pooling Layer. https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/
[13] Fully Connected Layers in Convolutional Neural Networks: The Complete Guide. (n.d.). https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/
[14] Team, K. (n.d.). Keras documentation: Layer activation functions. https://keras.io/api/layers/activations/
[15] Wang, C. (2019). The Vanishing Gradient Problem. https://towardsdatascience.com/the-vanishing-gradient-problem-69bf08b15484