Image Classification with Caffe Deep Learning Framework

Emine Cengil, Ahmet Çınar, Erdal Özbay
Computer Engineering Department
Firat University
Elazig, Turkey
{ecengil, acinar, erdalozbay }@firat.edu.tr

***Abstract*— Image classification is one of the important problems in the field of machine learning. Deep learning architectures are used in many machine learning applications such as image classification and object detection. The ability to manipulate large image clusters and implement them quickly makes deep learning a popular method in classifying images. This study points out the success of the convolutional neural networks which is the architecture of deep learning, in solving image classification problems. In the study, the convolutional neural network model of the winner of ilsvrc12 competition is implemented. The method distinguishes 1.2 million images with 1000 categories in success. The application is performed with the caffe library, and the image classification process is employed. In the application that uses the speed facility provided by GPU, the test operation is performed by using the images in Caltech-101 dataset.**

Keywords—image classification, deep learning, convolutional neural network, Caltech-101.

I. INTRODUCTION

Image processing is an improved example for digitizing a scene and performing some operations, or a method for extracting useful information from it. Image classification is a very wide range of image processing. Classification is the process of ensuring that unclassified images are included in their class within certain categories [1]. Image classification is a problem of computer vision that deals with a lot of basic information such as health, agriculture, meteorology and safety. The human brain can easily classify images. But for the computer this is not easy if the image contains noise. Different methods have been developed to perform the classification operation. General classification procedures can be divided into two broad categories as supervised classification based on the method used and unsupervised classification [2].

In a supervised class, the investigator defines homogeneous representations of information classes in the image. These examples are called training areas. The choice of appropriate training areas is based on knowledge of the analyst's classification. Thus, the analyst is tempted to control the classification of certain classes. The unsupervised classification reverses the supervised classification process. Programs used clustering algorithms are emphasize to determine statistical groupings or constructs in the data.

Generally, the analyst specifies how many groups or clusters in the data can be searched. In addition to specifying the required number of classes, the analyst can also determine the separation distance between the clusters and the parameters for the variation within each cluster. The unchecked classification does not start with a predefined class set. Supervised learning has been extremely successful in learning good visual presentations that not only give good results on the task they are trained on, but also transfer to other tasks and data sets. [3]. Academic and scientific areas have developed many methods for solving the image classification problem. These methods compete to perfection in image classification. Imagenet [4] is an image classification competition. The data to be processed every year and the number of categories to be classified are being increased. The competition, which was organized in 2012, has been a milestone in image classification.

A. Krizevsky et al. [5] have been successful in achieving the best result of recent times with the approach they have developed using convolutional networks. In the years that followed, the vast majority of those who participated in the competition used CNN. All of the entries have been developed using CNN. This contest proved the success of the CNN in image classification, which spoke much of its name. (CNNs) have become powerful models in feature learning and image classification. Researchers who have always aimed to achieve better have developed many methods using the CNN approach.

Michael et al. [6], point out that how to code information and invariance properties in deep CNN architectures is still a clear problem. In the study, it is suggested that CNN change the standard convolution block, and then transmit some more information layer after the blanket, but it is recommended to stay in the network with some stability. The basic idea is to take advantage of both positive and negative highs in convolution maps. This behavior is achieved by modifying the traditional activation function step before it is dumped. Comprehensive experiments in two classical datasets (MNIST and CIFAR-10) show that the proposed network performs better than standard CNN. H. Dou et al. [7] proposed a multi-scale CNN model with a depth-reducing multi-column structure to solve the problem of scale change in image

978-1-5386-0930-9/17/$31.00 ©2017 IEEE

classification. In particular, a coarse-grained pre-training method has been proposed to mimic human spatial frequency perception to train this multi-stage CNN, which accelerates the training process and reduces the classification error. In addition, model averaging techniques have been used to combine the models obtained during the preliminary training and to further improve performance.

Y. Zhou et al. [8], systematically examines the effect of image distortions on deep neural network (DNN) image classifiers. First, DNN classifier performance is examined under four kinds of degradation. Second, two approaches have been proposed to alleviate the effect of image distortion: retraining and fine-tuning with noisy images. Fine-tuning the findings with noisy images under certain conditions can alleviate much of the effect due to skewed entries and is more practical than re-education Y. Zhou et al. [9] presented a new method to classify medical images using a population of different convolutional neural networks (CNN) structures. They assume that different CNN architects learn to present semantic images at different levels, and thus better quality features of CNNs can be extracted. The method develops a new attribute extractor by fine-tuning the CNNs initiated in a large set of natural image data. The fine tuning process enhances the generic image properties from the natural images that form the basis of all images and optimizes them for various medical imaging methods.

Y. Zhou et al. [10] investigates the suitability and potential of the deep convulsive neural network in the controlled classification of Polari metric synthetic aperture radar (POLSAR) images. Designed with a two-stage convolutional layer, the deep neural network can automatically learn the hierarchical Polari metric spatial properties. X. D. Ren et al. [11] suggests that the CNN model has a problem of gradient diffusion that may cause slow updating of the basic parameters during the training process. To solve and improve the problem, this article presents a convolutional neural network model based on the introduction of basic component analysis for image classification. According to the image classification experiments on Mnist and Cifar-10 data sets, the model proposed in this article reduces the processes of iteration and optimization.

T. Williams et al. [12] uses convolutional neural networks (CNN) to classify handwritten digits in the MNIST database and images in the CIFAR-10 database. The proposed method preprocesses the field in the wavelet domain to obtain greater accuracy and comparable efficiency when compared to spatial domain processing. K. K. Pal, et al. [13] demonstrated the importance of preprocessing techniques for image classification using three variations of the CIFAR10 dataset and the Convolutional Neural Network. The results reported that Zero Component Analysis (ZCA) outperformed both the Mean Normalization and Standardization techniques for all three networks. Existing studies in the literature show the success of image classification of convolutional neural networks. However, the need for a more accurate, detail-oriented classification increases the need for changes, adaptations and innovations to deep learning algorithms. In this study, image classification is done using caffe deep learning library and Alexnet model.

The structure of this article is as follows: the second chapter presents the history, structure and properties of deep learning technology. The third section contains the proposed method is given, and the last section contains evaluation and future studies.

II. DEEP LEARNING

Machine learning, which began with probability and statistics in the past, had gone through some processes and has made significant progress. Today, the most advanced technique is the deep learning that has proved its success in many studies. In the 1950s, academic fields dealing with artificial intelligence introduced two methods for computer vision. Artificial Neural Networks and Decision Trees.

Artificial neural networks are the method inspired by the working mechanism of the human brain, and a layered network structure has been created by simulating the way the neurons work. This structure provided great successes from the end of the 1980s to the beginning of the 2000s [14]. Deep learning provides non-linear transformation of the data. Instead of shallow structures, it can model complex relations with gauss mixture models, hidden markov models, conditional random fields, multilayer structure. In Fig.1, deep learning neural network structure is given.

Deep learning algorithms first try to make a classification starting from the bottom of the picture. If classification cannot be done, it makes an abstraction to a top layer. In the picture, the pixels on the bottom layer are abstracted to a top layer, i.e. border lines. The new class that is created is passed to the top layer where the classification is made. Unlike other machine learning algorithms, deep learning does not involve any human intervention during these stages.

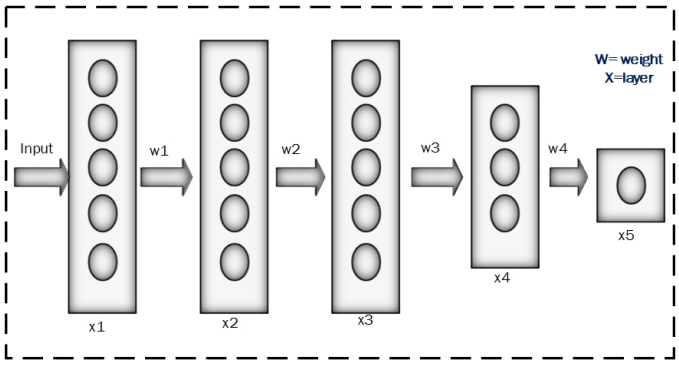

Fig.1. Network structure of deep learning

Every successful new layer, which the deep learning algorithm abstracts from the previous layer, increases the recognition and classification power of the deep learning algorithm. These layers are known as hidden layers. This feature makes the deep learning algorithm superior to other algorithms and allows for more classification and recognition with fewer examples taught in Algorithm.

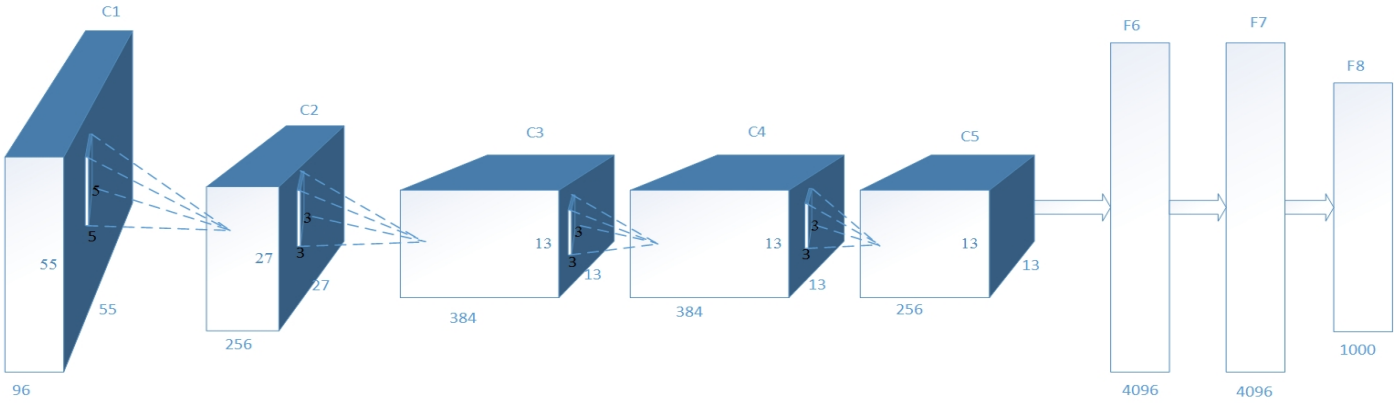

Fig.2. Network structure of deep learning

The most commonly used deep learning architects are; Automatic encoders, Restricted Boltzman Machines and convolutional neural networks [15].

III. PROPOSED METHOD

A. Convolutional Neural Network (CNN)

CNN is feed forward and is a very effective method of finding. The structure is simple; less training parameters and adaptability. Its weight-sharing network structure made it more similar to biological neural networks. This reduces the complexity and weight of the network model. CNN is used in many fields such as signal processing, natural language processing, robotics and sound processing in science and academia. But the most popular area is image processing and finding patterns.

Convolutional neural networks are very similar to ordinary neural networks. It consists of neurons with learnable weights and biases. Each neuron takes some inputs, generates a point product, and optionally follows it nonlinearly. The whole network expresses a differentiable scoring function. As in known neural networks, the neural networks of neurons include a loss function such as softmax in the last layer [16].

Convolutional networks differ from neural networks in terms of the form and function of layers. In ordinary neural networks, the layers are one-dimensional and the neurons in this layer are completely connected. On the other hand, the format of the CNN layers is usually three-dimensioned with width, height and depth parameters [17]. Fig. 2. shows the Alexnet architecture used for image classification.

Krishevsky's AlexNet consists of 11 layer. The first two layers are the convolution + max + norm, the third and fourth layers are convolution, the fifth layer is convolution + max, the sixth and seventh layers are fully connected and the final layer is the softmax..

The primary purpose of the convolution layer is to extract features from the input image. Convolution protects the spatial relationship between pixels by learning image properties using small squares of the input data. In the first layer, 96 filters of size 11x11 are used when convolution is done. The number of stride is 4, and the final image size of the convolution process is 55x55.

In the architecture of convolutional networks, it is common to add a pool layer periodically between one after another layers of convolution. In this section, there are also models where the sub-sampling layer replaces the pooling layer. The function is to gradually reduce the representative spatial dimension to reduce the amount of parameters and computations on the network and thus control over-fit. In the first pooling layer of the Alexnet model, 3x3 size filters are used. Image size after the process is 27x27. The same processes are repeated in subsequent layers.

The output at the end of the convolution and pooling layers represents the high level properties of the input image. There is no way of estimating the classification of the convolution and pool layers. The purpose of the Fully Connected layer is to use these properties to classify the input image into various classes based on the set of training data. Each neuron in the fully connected layer represents a class. There are 1000 neurons in the last layer because the method has been categorized for 1000 images.

B. Caffe (Convolutional architecture for fast feature embedding)

There are several powerful libraries that can be used to design and teach neural networks, including convolutional neural networks such as Theano, Lasagne, Keras, MXNet, Torch, and TensorFlow. Among them, Caffe is a library that can be used to research and develop real-world applications [18].

Caffe is a completely open-source library that gives open access to deep architectures. The library written in C ++ language is also implemented with matlab and python languages. Developed by Y. Jia, BVLC center. Caffe offers unit tests for accuracy, experimental rigor and installation speed, depending on the best practices of software engineering. In addition, the code is also well suited for research use because of its modularity and clean separation of the network definition from the actual implementation [18].

Models and optimizations can be done through the setting file without coding. The transition between CPU and GPU is seamless. Due to the high processing speed, it can be useful to use it in the image classification problem, which requires processing with millions of images. Caffe open source software is preferred because of these advantages.

C. Dataset

In practice, a pre-trained model using the ImageNet dataset is used. For testing, we use Caltech-101 [19] dataset. Database 101 contains images of categorized objects. The number of objects contained within the categories varies. In general, the number ranging from 40 to 800 in each category is around 50 in most categories. The size of the images is 300x200 pixels. Some images of the dataset are given in fig.3.

Fig. 3. Images of theCaltech-101 dataset [19]

D. Experimantal Results

The Imagenet model file contains model weights. In the Caffe library there is a model containing the weights of the model. The application was implemented using this model. In the Caltech-101 dataset, 30 categories were used for the test from 101 categorical images. Accordion, airplanes, ant, barrel, brain, butterfly, chair, dolphin, panda, cup, gramophone, laptop, scorpion, flamingo images are some of the categories we use.

The images of the ImageNet and Caltech-101 datasets are not the same as the way the images are categorized. While Imagenet classifies 1000 categories, CalTech-101 operates in 101 categories. However, some of the categories in the Caltech dataset (dollar_bill, faces, Garfield) are not found in Imagenet. In some images, a few images in Caltech can only be included in one category, whereas Imagenet contains more specific categories. For instance, if we think about the butterfly class, Caltech-101 has all butterflies in the butterfly class. However, Imagenet has identified five classes for butterfly, ringlet butterfly, monarch butterfly, cabbage butterfly, sulfur butterfly, and lichened butterfly. Table 1. Shows that the image information we use for the test operation and table 2 gives Imagenet dataset features used in the 2012 competition.

TABLE I. İMAGE CLASSIFICATION TEST RESULTS IN CALTECH-101 DATASET

Image size	Number of class	Number of images used	True Prediction	False Prediction
300x200	30	300	260	40

TABLE II. THE DATASET USED FOR TRAINING THE MODEL: PRE-TRAINED MODEL

Image Size	Number of class	Number of images
flexible	1000	1.2 million

Fig.4. shows the classes predicted correctly. From the images we gave, beaver, scorpion, strawberry flamingo and accordion classes are common classes for both dataset. The others are divided into more specific classes in the Imagenet database.

Fig. 4. Correctly classified objects

Fig. 5 gives incorrectly classified images and table 3 gives the classes should be the class is incorrectly predicted.

TABLE III. TRUE CLASSES OF MISCLASSIFIED IMAGES

True class	Predicted class	Class Label
brain	Golf_ball	n03445777
bonsai	Aircraft carrier	n02687172
cellphone	mousetrap	n03794056
headphone	Bolo tie	n02865351
gramophone	Electric fan	n03271574
elephant	hammerhead	n01494475

Fig. 5. False classified objects

Since the yield categories are not identical, it is not possible to give a clear accuracy rate. However, in general, it is seen that the accuracy of the correct class estimate is higher for each class.

IV. Conclusions

It was performed in the caffe library using the Alexnet model, which was constructed using convolutional neural networks. Image classification was done using a convolutional neural network model that did thousands of categorization of images. There are many publicly available datasets for image classification. Caltech-101, one of these, was used to see the performance of the model. 300 images are selected from 30 categories obtained by the dataset. According to the categories included in Imagenet, high accuracy is obtained in finding images of the same classes. In addition to this, we have also found other classes more specifically. Some classes in Caltech, because the ImageNet was not introduced, the classes of those images were found incorrect.

Image classification is done as a result of our application using python and caffe libraries in the ubuntu operating system. Although the method tested with Caltech-101 data yields accurate results in general, we cannot give a clear accuracy since the labels after the training and the classes of the test data are not exactly the same. The method that was tested with Caltech-101 data yielded accurate results in general. Since the tags created after training and the classes of test data are not exactly the same, we cannot give a clear accuracy.

References

[1] M. Sonka, V. Hlavac, and R. Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014.

[2] K. B. Johnston, and H. M. Oluseyi. Generation of a supervised classification algorithm for time-series variable stars with an application to the LINEAR dataset. *New Astronomy*, *52*, 35-47, 2017.

[3] N. Srivastava, E. Mansimov, R. Salakhudinov. Unsupervised learning of video representations using lstms. In: International Conference on Machine Learning. 2015. p. 843-852.

[4] O. Russakovsky, J. Deng, H. Su, J., Krause, S. Satheesh, S. Ma,... and A. C. Berg,. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252, 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton ImageNet classification with deep convolutional neural networks. In NIPS, pp. 1106–1114, 2012.

[6] M. Blot, M. Cord, N. Thome. Max-min convolutional neural networks for image classification. In: Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016. p. 3678-3682.

[7] H. Dou, and X. Wu. Coarse-to-fine trained multi-scale convolutional neural networks for image classification. In Neural Networks (IJCNN), 2015 International Joint Conference on pp. 1-7, 2015, July.

[8] Y. Zhou, S. Song, and N. M. Cheung. On Classification of Distorted Images with Deep Convolutional Neural Networks. *arXiv preprint arXiv:1701.01924*.2017

[9] A.Kumar, J. Kim, D. Lyndon, M. Fulham and D.Feng, An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE journal of biomedical and health informatics, 2017, 21.1: 31-40.

[10] Y. Zhou, H. Wang, F. Xu, and Y. Q. Jin, Polarimetric SAR Image Classification Using Deep Convolutional Neural Networks. IEEE Geoscience and Remote Sensing Letters, 13(12), 1935-1939.2016

[11] X. D. Ren, H. N. Guo, G. C. He, X. Xu, C. Di, and S. H. Li. Convolutional Neural Network Based on Principal Component Analysis Initialization for Image Classification. In Data Science in Cyberspace (DSC), IEEE International Conference on (pp. 329-334). IEEE.2016.

[12] T. Williams, and R. Li. Advanced Image Classification Using Wavelets and Convolutional Neural Networks. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on pp. 233-239, 2016.

[13] K. K., Pal, and K. S. Sudeep. Preprocessing for image classification by convolutional neural networks. In Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on pp. 1778-1781, 2016.

[14] G. Doğan, , Yapay Sinir Ağları Kullanılarak Türkiye'deki Özel Bir Sigorta Şirketinde Portföy Değerlendirmesi, Master thesis, Hacettepe University, Statistics Department, 2010

[15] E. Cengil, A. Cinar, A New Approach for Image Classification: Convolutional Neural Network, European Journal of Technic (EJT), 6(2), 96-103, 2016.

[16] http://cs231n.github.io/convolutional-networks/

[17] A. Karpathy, Convolutional Neural Networks (CNNs / ConvNets). CS231n Convolutional Neural Networks for Visual Recognition, August 7, 2016.

[18] H. H. Aghdam, E. J. Heravi. https://link.springer.com/chapter/10.1007/978-3-319-57550-6_4, 2017.

[19] L. Fei-Fei, R. Fergus and P. Perona. *Learning generative visual modelsfrom few training examples: an incremental Bayesian approach tested on 101 object categories*. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004