International Journal of Pure and Applied Mathematics
Volume 119 No. 17 2018, 1307-1319
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue
http://www.acadpubl.eu/hub/

Image Classification using Convolutional Neural Networks

Muthukrishnan Ramprasath
Sr. Assistant professor, Department of Computer Science & Engineering,
Madanapalle Institute of Technology & Science,Andhra Pradesh, INDIA
mramprasath@mail.com
M.Vijay Anand
Professor, Department of Computer Science and Engineering
Saveetha Engineering College, Chennai, India
mvijay200304@yahoo.co.in
Shanmugasundaram Hariharan
Professor, Department of Computer Science and Engineering,
Saveetha Engineering College, Chennai, India
mailtos.hariharan@gmail.com

Abstract:

In recent year, with the speedy development in the digital contents identification, automatic classification of the images became most challenging task in the fields of computer vision. Automatic understanding and analysing of images by system is difficult as compared to human visions. Several research have been done to overcome problem in existing classification system, but the output was narrowed only to low level image primitives. However, those approach lack with accurate classification of images. In this paper, our system uses deep learning algorithm to achieve the expected results in the area like computer visions. Our system present Convolutional Neural Network (CNN), a machine learning algorithm being used for automatic classification the images. Our system uses the Digit of MNIST data set as a bench mark for classification of grayscale images. The grayscale images in the data set used for training which require more computational power for classification of images. By training the images using CNN network we obtain the 98% accuracy result in the experimental part it shows that our model achieves the high accuracy in classification of images.

Keywords: Deep learning, Convolutional Neural Network (CNN), image classification, MINIST Image Datasets, Machine Learning.

1. Introduction

In recent years, due to the explosive growth of digital content, automatic classification of images has become one of the most critical challenges in visual information indexing and retrieval systems. Computer vision is an interdisciplinary and subfield of artificial intelligence that aims to give similar capability of human to computer for understanding information from the images. Several research efforts were made to overcome these problems, but these methods consider the low-level features of image primitives. Focusing on low-level image features will not help to process the images.

Image classification is a big problem in computer vision for the decades. In case of humans the image understanding, and classification is done very easy task, but in case of computers it is very expensive task. In general, each image is composed of set of pixels and each pixel is represented with different values. Henceforth to store an image the computer must need more spaces for store data. To classify images, it must perform higher number of calculations. For this it requires systems with higher configuration and more computing power. In real time to take decisions basing on the input is not possible because it takes more time for performing these many computations to provide result.

In [1], has discussed extraction of the features from Hyper Spectral Images (HSI) by using Convolutional Neural Network (CNN) deep learning concept. its uses the different pooling layer in CNN for extraction of the feature (nonlinear, Invariant) from the HIS which are useful for perfect classification of images and target detection. It also addresses the general issues between the HSI images features. In the perspective of engineering, it seeks to automate tasks that the human visual system can do. It is concerned with the automatic image extraction, analysis and understanding useful information with images.

In last decade, several approaches for image classification was described and compared with other approaches. But in general image classification refers to task of extracting information from the image by labelling the pixels of the image to different classes. It can be done in two ways one is Supervised classification, Unsupervised classification. In [2], has discuss the use of the Unsupervised learning algorithm in underwater fish recognition framework for classifying images.

This technique the pixels of the image are clustered into groups without intervention of the analyst. Grounding on the clustered pixels the information is retrieved from the image. In real world the availability of labelled data is very less hence unsupervised classification is done in most cases. In [3], has discussed Supervised classification techniques that analysis and train the classifier on the labelled images and extracting features from them. By using the learned information of the training, the newly provided image will be classified basing on the features observed in the image.

Now a day, Deep learning algorithms are providing successful results in the areas like computer vision. The Convolutional Neural Network, a machine learning algorithm is being used for the image classification. In [4], uses deep learning algorithm for classify the quality of wood board by using extracted texture information from the wood images. he also made the comparison with machine learning architecture. CNN is a type of feed-forward artificial neural network that has been successfully applied to analyse visual images. It is inspired by the biological processes and the neurons are connected as in animal visual cortex. In [5], has discussed automatic recognition cattle images using CNN which helps to extract the necessary characteristic from the cattle images and Support Vector Machine (SVM) techniques is used for classification of those images.

In [6], has uses high resolution images in ImageNet data set having 15 million labelled images with 1000 different classes used for classification with help of deep convolutional Neural Network. CNN having three different layers such as input layer, hidden layers and an output layer. In general images is constructed as a matrix of pixels and these pixel values are given as input to input layer along with weights and biases (for non-linearity). The output layer will be a fully connected layer usually to classify the image to which class it belongs to. The hidden layer may be convolutional, pooling or fully connected. In [7], has discuss the manifold -learning techniques for classifying remotely sensed hyperspectral data.

The Convolutional layer is core building block and has learnable filters as parameters. Each filter is spatially small (width and height) but extends across the depth of the input volume. The 2-dimensional activation map is produced by performing dot product between input and entries of filter for every filter. As a result, the network learns filter that activate when it detects specific feature in some spatial position in the input. The pooling layer is used in down sampling the image without losing any information from the image.

Max pooling uses the maximum value from the cluster of neurons at prior layer. The fully connected layer connects every neuron in one layer to every neuron in other layer. CNNs use little pre-processing when compared to traditional classification algorithms which use filters that are hand engineered. The independence of human intervention in learning filters is good advantage of CNN.CNN is supervised deep learning approach which requires large labelled data for training on the network. After training the model will learn the weights and the accuracy of the classifier is improved.

Then an image is given as input and the classifier shows to which class it belongs to. Google's self-driving car is a novel deep learning project from Google company is an example for the recent development in the field of artificial intelligence. For this project the image data is provided as input from the real world and the decisions are made basing on the information gained from the image. Here the image classification is done, and the decisions are taken basing on it. If the image has road the car will go straight, if there is any obstacle like tree or human then the car is stopped. Facial recognition project from Facebook in which the photo of the user is identified by deep learning. The network is trained over some human faces and features from face like eyes, ears and nose are learnt from the training of the network. The classifier classifies the face based on the features observed in the images. In [8], has discussed reorganization of different species of animal and plant using Unsupervised learning algorithm. This reorganization process mainly focuses the similarity in shape and structure of the species shared across different categories and difference in the object parts. He also proposes the template model for capturing the common part and shapes of the object parts and Co-occurrence relation of the object pattern.

2. Related work

Image classification is a widespread research area in the field of deep learning, Pattern recognition, Human Computer Interaction and got substantial attraction in many research scientists. In [9], classification of images done by extracting the features from the image. Usually most midlevel feature learning methods focus on the process of coding or pooling but here they emphasize that the mechanism behind the image composition also strongly influences the extraction of features from the image.

In feature extraction, image content exploration is effectively done by using hierarchical image decomposition approach. Here each image is decomposed into a series of semantic components like the structure and texture images. The semantic image content (structure and texture) can be matched with other images by using various feature extraction methods. The following two different schemas used to for representation of different image property related feature such as Hand-crafted features used in single staged network and the second ones learns features from raw pixels automatically by multistage network.

In [10], has discussed classification of Natural images using biological stimulated model. Its uses well known analogous progress in visual information system and inference procedure of human brain functionality. This model primarily used for image analysis and Natural classification. This system is composed of three important units as biologically inspired visual selective attention unit, Knowledge structuring unit and Clustering of visual information unit. It uses the low-level features in the images to automatic extraction of important relationships between images. The system follows the limitation in the human visual system to achieve higher accuracy in classification of images.

The biologically inspires system having two components namely Bottom-up saliency map module which produces a salient area from the low-level features extracted from natural images and Top-down selective attention module which performs decisions on interesting objects by interaction of humans. These two components closely follow the mechanisms of the visual what pathway and where pathway of the human brain. These components have been integrated in knowledge structuring unit., The clustering of visual information is achieved by using output of the knowledge clustering unit and it is based on high-level top-down visual information perception and classification model.

In [11] has discussed spoof finger print detection using Convolutional Neural Network (CNN), the goal of biometrics is to discriminate automatically between subjects in a reliable way and as per target applications basing on one or more signals derived from traits like face, fingerprint, iris, voice or written signature. There are more advantages from biometric technology than the traditional security methods based on something we remember or know like PIN, PASSWORD and something physically we have like KEY, CARD etc.

In yearly days Several [12,13,14] fingerprint detection algorithm has been proposed and they can be divided into two categories the namely Hardware and Software. In hardware approaches specific device is attached to any hardware sensor device to detect the living attribute in the Human such as blood pressure, heart beat rate etc.

Finger print image feature are used to distinguish between real and fake fingerprints. In this model two feature extractors have been used namely Convolution Networks and Local Binary Patterns. In the interim Support vector machine classifier (SVM) also used in conjunction with both techniques for classification of original and fake finger prints. This system uses the dataset comprising of real and fake fingerprints images retrieved from different sensors to train the model. Fingerprints are obtained from following sensors namely Biometrika FX2000, Digital 4000B, Italdata ET10 and fake fingerprints were obtained from different materials like gelatin, wood glue, eco flex and silgum. To training the classifier the following four different phases used such as,

1. Pre-processing of data using image reduction, contrast equalization, filtering and region of interest extraction.
2. Feature extraction done by implementing two techniques LBP and Convolutional Network
3. Data normalization and dimensionality reduction.
4. Classification of the images using SVM.

Extraction of patterns found within local regions of the input images done by convolutional networks that are common throughout dataset. Local Binary patterns (LBP) is used for feature extraction in texture descriptors normally. Using both the methods feature extraction is done and in pipeline these are applied distinctly.

Human face detection is becoming a very important in the fields of image reorganization due to extensive growth related to its applications in various fields like security access control, advanced human-computer interaction and content-based video indexing etc. In [14], has discuss novel detection approach for detection complex types of face image having variable image pattern in the real world. It's also uses the CNN for automatic extraction of image feature from the training set having set of face and Non-face images. All the existing algorithm uses local facial features for detecting human faces

In last decade many approaches [15,16] has been presented and compared about local facial features and classification of images using geometric model of human face. Some other approaches focus on template matching methods used to detect the local sub feature of images. This system detects semi-frontal human faces in complex images datasets by classifying the image as face image or non-face image. CNN also used to automatically extract the important feature from the images. Furthermore, the pre-processing of image is not required, and fast processing is automatically done by successive simple convolutional and subsampling operations.

CNN having three different kinds of layers to process the images. They are convolution layers, sub sampling layers and classification layers containing sigmoid neurons. The convolutional layers containing certain number of planes. Each plane is considered as a feature detector. After finding the image feature then locate the images. Then subsampling layer used to performs input dimension reduction by preserving the information of the image. Then it uses sigmoid neurons function to perform the classification operation.

The rest of the paper is organised as follows section 3 present the proposed system architecture for image classification using CNN, section 4 implementation section 5 present conclusion and future enhancement.

3. Proposed system architecture for image classification

Computer vision is an interdisciplinary field of machine learning and artificial intelligence and is concerned with the automatic extraction, analysis and understanding useful information from images. With recent advancements of technology there is explosive growth in digital content regarding images and videos. In the field of computer vision understanding and analysing the images is a crucial problem by the computer as compared to human. So, the classification of images will be done with help of human intervention. The human uses the Realtime images datasets (MNIST digit images) for training and testing purpose.

The grayscale images form MNIST data set given as input

Initially, the human will train classifier to obtain the desired pattern from the images. Then the images classified with help of the pattern precisely obtained from the previous stages. The obtained results will vary with respect to the patterns observed and it is completely dependent on the knowledge of the person who classifies. In [17,18], has discussed deep learning architecture for classification of images also it uses different layers in Convolutional Neural Network (CNN) to extract new feature form the images datasets. The figure 1 explain the components of CNN networks.

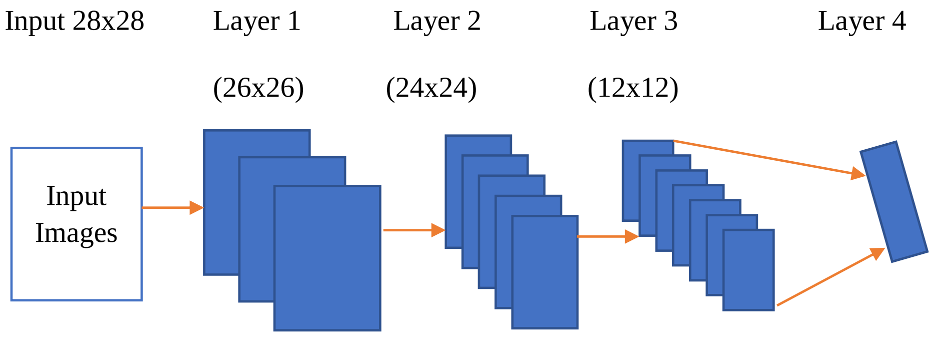

Figure 1: Architecture of Convolutional Neural Network (CNN)

Our system uses grayscale images as input image having 28x28 sizes. The first layer in CCN applied 32 filters on input images, each image size is 3x3 producing 32 feature maps of size 26x26. The second layer is applying 64 filters, each of size 3x3 producing 64 feature maps of size 24x24. Max pooling layer is act as third layer which is used to down sampling the images to 12x12 by using subsampling window of size 2x2. The layer 4 is fully connected layer having 128 neurons and uses sigmoid activation function for classification of images and produce the output image. The figure 2 show the CNN architecture and its components.

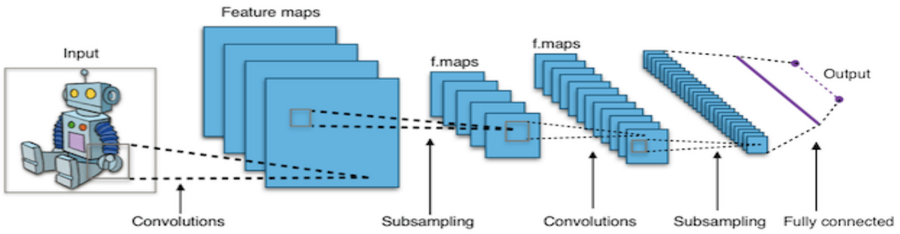

Figure. 2 Typical CNN Architecture

In regular neural networks (feed-forward), each hidden layer is made up of a set of neurons and each neuron is connected to the hidden layer. The last layer of network is fully connected and used for classifying images. In general, the size of the image is 28*28*1 (28 wide, 28 high, 1 colour channel) provided as input, then the first hidden layer would have 28x28x1=784 weights. This amount of weights seems still manageable. In case of larger images with size 400x400x3, requires 400*400*3= 4,80,000 weights, this fully connected layer does not scale much.

Convolutional neural networks will have the distinct architecture than regular neural networks with that advantage it takes input as different size of images. The layers of a convnet have neurons arranged in three dimensions width, height and depth. It is to be noted that the word depth is referring to the third dimension of an activation volume, not to the depth of a full Neural Network, which can refer to the total number of layers in a network. Let consider input images of size 32x32x3 and the volume has dimensions 32x32x3 (width, height, depth respectively).

3.1 Layer used for Building Conv Nets

Convnet is a sequence of layers and every layer of convnet transforms one volume of activations to another using a differentiable function. Most important types of layers are used to build ConvNet architecture which will be called as Convolutional layer, Pooling layer and Fully connected layer. These layers will be stacked to form a full ConvNet architecture. Input image will give as input to the layer it holds the raw pixels values of the input images.

Conv layer compute the output of neurons which are connected to local regions in the input and each neuron performs the computation of dot product between their weights and a small region they are connected to in the input volume. RELU layer leaves the input volume unchanged as such if 28x28x1 is given as input volume, the output volume would be 28x28x1. It will apply an elementwise activation function like max (0, x) thresholding at zero.

In pooling layer dimensions of the image will be reduced but the information of the image is retained. The down sampling operation is performed along the spatial dimensions (width, height) that is if the input is 24x24x64 then the output volume would be 12x12x64. Finally, fully connected layer computes the class scores resulting in volume of size 1x1x10 where each of the 10 numbers correspond to a class score such as among the 10 categories given as input. convolutional neural networks will transform the original image layer by layer from the original pixel values to the final class scores. It is important to consider that some layers will have parameters and some layers don't have. The convolution and fully connected layers perform transformations that are a function of not only the activations in the input volume, but also of the parameters (the weights and biases of the neurons) and the RELU/POOL layers will implement a fixed function.

4. IMPLEMENTATION OF PROPOSED SYSTEM

Our proposed system uses CNN for implementation purpose. Convolutional Neural Networks are very similar to ordinary Neural Networks, that are made up of neurons that have learnable weights and biases. Every neuron performs dot product by receiving some input and using bias it follows non-linearity. The whole convent still expresses a distinct score function, from the raw pixels on one end to class scores at the other end.

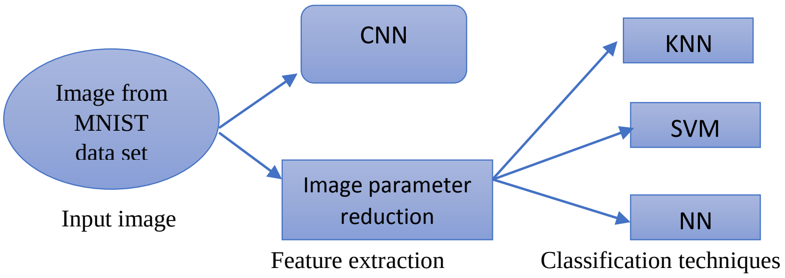
CNN
Image from MNIST data set
Input image
Image parameter reduction
Feature extraction
KNN
SVM
NN
Classification techniques

Figure 3. Image classification using machine learning techniques.

They have a loss function like SoftMax on the last layer which is fully connected layer. As the inputs are images to convent, it allows to encode certain properties in architecture. These properties make the forward function more efficient to implement and vastly reduce the number of parameters in the network. The mail goal of the image classification able to extract the feature from raw images

Algorithm

1. Batch size =128 , no of classes 10, number of epochs = 5,

2. Dimension of input image 28 ×28,
3. Loading the input images from MNIST data set
4. Variable exploration: X=test data set (10000,28,28,1), Train data set (60000,28,28,1)
5. Creating and compiling the models
6. Training the network.

The above algorithm explains the general steps involved in training and testing the MNIST data set for image classification in CNN.

In General networks each neuron is connected to all neurons of previous layer. In real time it is impractical for high-dimensional inputs such as images. For instance, the input volume has size 32x32x3 and the receptive field size is 5x5.

EPOCH	LOSS	ACC	VAL_LOSS	VAL_ACC
1/5	0.3450	0.8955	0.0843	0.9739
2/5	0.3452	0.8955	0.08431	0.9617
3/5	0.0448	0.0875	0.0874	0.9743
4/5	0.0451	0.9854	0.0729	0.9787
5/5	0.0628	0.9811	0.0444	0.9860
Total loss	0.5412	0.9842	0.04438	0.986

Table 1. The loss and accuracy of all epochs

Then each neuron of conv layer will have weights to a 5x5x3 region in the input volume for a total of 5*5*3=75 weights (and +1 bias parameter). The extent of connectivity along the depth axis must be 3, since it is the depth of the input volume. In CNN parameter sharing is used to reduces the number parameter in entire process.

Table 1 shows the losses and accuracy of each epochs

5. CONCLUSION

In this paper, we used Convolutional Neural Networks (CNN) for image classification using images form hand written MNIST data sets. This data sets used both and training and testing purpose using CNN. It provides the accuracy rate 98%. Images used in the training purpose are small and Grayscale images. The computational time for processing these images is very high as compare to other normal JPEG images. Stacking the model with more layers and training the network with more image data using clusters of GPUs will provide more accurate results of classification of images. The future enhancement will focus on classifying the colored images of large size and its very useful for image segmentation process.

References

1. Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, “Deep feature extraction and classification of Hyperspectral images based on Convolutional Neural Network (CNN)” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 10, pp. 6232-6251, 2016.
2. Meng-Che Chuang, Jenq-Neng Hwang, Kresimir Williams, “A Feature Learning and object Recognition Framework for Underwater Fish Images”, IEEE Transactions on Image Processing, vol. 25, no. 4, pp. 1862-72, 2016.
3. Meng-Che Chuang, Jenq-Neng Hwang, Kresimir Williams, “Supervised and Unsupervised Feature Extraction Methods for Underwater Fish Species Recognition”, IEEE Conference Publications, pp. 33-40, 2014.
4. Hanguen Kim, Jungmo Koo, Donghoonkim, Sungwoo Jung, Jae-Uk Shin, Serin Lee, Hyun Myung, “Image-Based Monitoring of Jellyfish Using Deep Learning Architecture”, IEEE sensors journal, vol. 16, no. 8, 2016.
5. Carlos Silva, Daniel Welfer, Francisco Paulo Gioda, Claudia Dornelles,” Cattle Brand Recognition using Convolutional Neural Network and Support Vector Machines ”, IEEE Latin America Transactions, vol. 15, no. 2, pp. 310-316, 2017.
6. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, 2012, pp. 1106–1114.
7. D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, “Manifold-learningbased feature extraction for classification of hyperspectral data: review of advances in manifold learning,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 55–66, Jan. 2014.
8. S. Yang, L. Bo, J. Wang, and L. G. Shapiro, "Unsupervised template learning for fine-grained object recognition," in Advances in Neural Information Processing Systems, 2012, pp. 3122-3130.
9. Jianjun Qian, Jian Yang, Yong Xu, “Local Structure-Based Image Decomposition for Feature Extraction With Applications to Face Recognition”, IEEE transactions on image processing, vol. 22, no. 9, September 2013 pp.3591-3603.
10. Le Dong and Ebroul Izquierdo “A Biologically Inspired System for Classification of Natural Images” IEEE transactions on circuits and systems for video technology, vol. 17, no. 5, may 2007, pp. 590-603.

11. Y. Chen, A. Jain, and S. Dass, “Fingerprint deformation for spoof detection,” in Biometric Symposium, 2005, p. 21.

12. B. Tan and S. Schuckers, “Comparison of ridge-and intensity-based perspiration liveness detection methods in fingerprint scanners,” in Defense and Security Symposium. International Society for Optics and Photonics, 2006, pp. 62 020A–62 020A.

13. P. Coli, G. L. Marcialis, and F. Roli, “Fingerprint silicon replicas: static and dynamic features for vitality detection using an optical capture device,” International Journal of Image and Graphics, vol. 8, no. 04, pp. 495–512, 2008.

14. C. Garcia, M. Delakis, “Convolutional face finder: a neural architecture for fast and robust face detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 11, pp. 1408-23, 2004.

15. M. Yang, D. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002.

16. E. Hjelma˚s and B.K. Low, “Face Detection: A Survey,” Computer Vision and Image Understanding, vol. 83, pp. 236-274, 2001.

17. A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Neura Inf. Process. Syst.*, Lake Tahoe, NV, USA, 2012, pp. 1106–1114.

18. G. Hinton and R. Salakhutdinov, “Reducing the dimensionality of datawith neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.