Image Classification Using Convolutional Neural Network

Savitha G
Department of Computer Science and Engineering
RV Institute of Technology and Management
Bengaluru, India

Mahalakshmi L
Department of Computer Science and Engineering
RV Institute of Technology and Management
Bengaluru, India

Kusuma L
Department of Computer Science and Engineering
RV Institute of Technology and Management
Bengaluru, India

***Abstract*–Image classification is essential to identify various images and categorize them for different purposes. Convolutional neural network architecture is a class of neural networks that has multiple layers to extract features. This architecture is mainly used for identifying and recognizing objects in an image. In this paper, we identified different CNN architectures for classifying images like tree species, hyperspectral images, medical images, aerial images, food images, and plankton images. Our study shows that CNN has high computational efficiency and decreases image's high dimensionality without information loss.**

Keywords—Convolutional Neural Network, Image Classification, Extract features, multiple layers, High dimensionality

I. INTRODUCTION

Image Classification is a technique to extract features from an image to observe the patterns in the dataset. Image classification is used in numerous applications like agriculture, surveillance, object detection, medical and has become an interesting field of research. Image classification using Convolutional neural networks reduces the need for pre-processing and brings down the error rate for classifying the images.

The Convolutional Neural Network (CNN) is a multi-layer neural network that extracts features from images and classifies them based on these features. CNN has many layers for extracting features and generating feature maps. Feature maps are the arrays that represent the input and output of each stage. The layers of CNN are the Convolutional layer, Non-linearity layer, Pooling layer, Fully connected layer, and Dropout layer. The convolutional layer consists of filters (or kernels) that convolve throughout the image and produce the feature maps. Non-linearity layer uses these feature maps as input and applies its activation functions like the Linear function, Sigmoid function, Tanh function, Rectified Linear (ReLU) function, and Softmax function to generate activation maps. Then the pooling layer uses any one of the pooling techniques to reduce the dimensionality of the activation maps. The Fully connected layer connects the neurons of one layer to another. The Dropout layer is used to reduce the impact of overfitting. CNN architecture has progressive stages of downsampling and upsampling. Downsampling is used to produce an abstract representation of the input image like changing its resolution to extract its features. In upsampling, spatial dimensions of the abstract image are made equal to the input image, for example retaining its original resolution.

II. LITERATURE SURVEY

Zezhong Zheng et al., [1] proposed a deep convolutional neural network to extract features from hyperspectral images and classify them using logistic regression classifiers. Hyperspectral images are images that collect and process information across the electromagnetic spectrum. Datasets were gathered using reflective optic system imaging spectrometer sensors. Principal Components Analysis is used to extract features from images and reduces the correlation and dimension of the feature. CNN model is trained and then tested with respective datasets. Experimental results show that CNN architecture had an overall accuracy of 93.39%; hence, it was a robust classification approach for Hyperspectral images.

Asia Kausar et al., [2] proposed a Pure-CNN(PCNN) framework to classify fruit images from the Fruit-360 dataset with 55244 color images in 81 categories. This architecture has Input, convolution, strides, ReLU, Global Average Pooling (GAP), and Softmax layers as the parameters in the training phase. During testing, a forward pass is applied for high-score prediction. PCNN with GAP has the highest accuracy of 98.88% and 46 wrongly predicted images. This method using GAP reduces overfitting and hence is very efficient.

Guoqing Ding et al., [3] proposed a Fully trainableMulti-layer CNN model to classify fish images collectedfrom live video datasets from live video having images of 4 different species. CNN captures high nonlinear mapping between input and output. The model is trainedusing Regularization, Batch approach, and Backpropagation. A learning rate of 0.3 and a Batch sizeof 20 is used. 3 different CNN Models were designed and model 2 had the highest accuracy of 96.51% and Extreme Learning Machine (ELM) of 85.52%.

Aws Anaz et al., [4] presented various classifiers for classifying hand poses from datasets having 8 poses in 17 subjects. Designed CNN and Alexnet were the two CNN models proposed along with Support Vector Machine (SVM). Designed CNN used ReLU activation and Softmax functions to predict the class. Alexnet has convolution and FC layers and is pre-trained with the classes of these datasets. Among all the methods, Alexnet has the highest accuracy of 99% and compute time of 8ms per sample.

Hui Li et al., [5] presented the tree species classification using airborne lidar data and high-resolution multispectral (Worldview-2) images. CNNs models evaluate for tree species classification at individual tree level. These classifications are based on a combination of Lidar data and high-resolution multi-spectral images. When the CNNs models (ResNet-18 and DenseNet-40) were evaluated with RF and SVM the ResNet-18 and DenseNet-40 gives an accuracy of 0.889 and 0.880.

Le Kang et al., [6] presented the Document image classification using a Convolutional neural network. Initially, the original images will have high resolution to reduce this it is downsampled and resized to 150x150. The network architecture is $150\times150-36\times36\times20$ $-8\times8\times50-1000-1000-M$ where M is the number of classes. The data sets were used here Tobacco litigation dataset and the Netaji Subhas Institute of Technology (NSIT) tax-form dataset. When the CNN is evaluated for these data sets it obtained the highest accuracy of 65.35%.

Meet P. Shah et al., [7] presented the leaf classification using marginalized shape context, shape, and texture dual-path convolutional neural network. The dual-path CNN is used to find joint feature presentations for leaf images and use their texture and shape features. It also optimizes the features for further classification processes. The datasets used here are Flavia, leafsnap, and Imageclef. This work gives the best results when this approach is evaluated against vanilla CNN and some handcrafted shape features.

Pantelis I. Kaplanoglou et al., [8] presented the Margin based sample filtering for image classification using convolutional neural networks. This work aims to describe the effective way to speed up the training procedures of CNNs. This work proposed the sample filtering idea (Algorithm) which reduces the cost of training to filter the samples. this sample filtering root on the multi-class margin. The datasets used in this work are Less ImageNet Training Examples (LITE) datasets obtainable on the ILSVRC2012 training set. This method speeds up the CNNs for image classification by 18% with a slight drop in performance benchmark by around 1%.

Ping Han et al., [9] presented the polarimetric synthetic aperture radar image classification. This method makes use of the optimal feature selection and convolutional neural network. In this work classification algorithm is proposed based on feature selection with thecombination of CNN. Lee filtering was performed on the Polarimetric Synthetic Aperture Radar (PolSAR) imageand computed optimized features of the image. Then thePolSAR image is divided into a sub-feature image with size nxn and every block consisting of 6 feature imagesand then put for training and then classification results are obtained for the whole image. This method gives theimproving final accuracy and verifies the better classification performance. The drawback of this work requires a manually set of appropriate parameters whichextend the complexity of the experimental procedures.

III. CLASSIFICATION OF PLANKTON IMAGES

Yilong Zhang et al., [10] proposed a CNN model to classify marine plankton with digital holography. A Dataset consisting of macroscopic images of marine plankton is obtained from the Kaggle website. Original objects are extracted from unreconstructed holograms. The data transfer learning neural network model used with ResNet mode (DTL-ResNet18) is used to extract features and classify the raw plankton images. The model is pre-trained with color images in ImageNet. The loss function during training is improvised using the Adam optimization method. This model takes less time to classify as it does not involve the reconstruction phase for images. It has an accuracy of 94.59% for object images and 92% for holograms. Hence it is efficient in classifying raw plankton images.

Angang Du et al., [11] presented plankton image recognition using deep convolutional neural networks along with second-order pooling layers. This deep CNNs architecture contains many parts such as a convolutional layer, Batch normalization, ReLu, Matrix Power Normalized Covariance (MPN-COV) pooling layer, and a fully connected layer. In the first convolutional layer, they used a kernel of 7x7 and for others, it was 4x4. Normalization is used to speed up the training procedure and perfection by managing the distribution covering layers. Relu is used as an activation function to the output of each convolutional layer. MPN-COV pooling layer is used to get a normalized covariance matrix and then a fully connected layer drives the last classification resolution. The dataset used here is PlanktonSet 1.0. These models MPN-COV-ResNet-50 and MPN- COV-ResNet-101 give the highest accuracy of 75.63% and 75.82%.

IV. CLASSIFICATION OF FOOD IMAGES

David et al., [12] proposed a CNN architecture having 2D convolutional layers and web scraping for classifying food images to build a diet monitoring system. A food-101 dataset was used with 101000 images in 101 categories. Image preprocessing is done using various parameters and a Google inception V3 model pre-trained on imageNet is retained. Pooling, dense, and softmax functions are used to extract features, define dimensions and identify classes respectively. Stochastic gradient descent is used to increase performance and the learning rate scheduler uses epochs as input with output as the new learning rate. This architecture has an accuracy of 86.97%. CNN is more appropriate for large datasets with more classes.

K. Kogias et al., [13] proposed a two-level CNN classification system to classify the NTVA-Food 2017 dataset having 3248 images in 8 different categories. CNN has convolution, non-linearity, and pooling layers for extracting features and fully connected layers to classify images. In the first level, the images are classified into eight broad categories and in the second level assign each image to one of these categories. This method attained an accuracy of 85.94% for second-level classification.

V. CLASSIFICATION OF AERIAL IMAGES

Yuansheng Hua et al., [14] presented the end-to-end convolutional neural network and this method is practiced to fuse multi-level characteristics for aerial classification. This network contains a single convolutional stage and an atrous(dark) convolution stage and these are used to produce low-level and high-level features. The dataset used here is the UC-Merced dataset (UCM) which contains 2100 images of 256x256. These are categorized into 21 land-use classes Aerial Image Dataset (AID) contains images of 30 aerial scene classes. In this work, the evaluated methods are BoVW, VGG-VD-16, GoogLeNet, and LAHNet. The LAHNet gave the highest accuracy 99.10% for UCM and 95.78% for AID among all these.

Tianyu Tang et al., [15] presented the coupled region-based convolutional neural networks(R-CNNs). This model detects the vehicles automatically in aerial images. There are two networks proposed here namely the Vehicle proposal network (VPN) and the Vehicle classification network (VCN). VPN was presented to identify the candidate vehicle-like places by using the hyper feature map combined with distinct layers of feature maps.VCN is developed for the verification of candidate regions and also the classification of vehicles in eight directions. Two datasets are used in this work. The Munich Vehicle dataset consists of 20 aerial images with a ground sampling distance of 13cm. Another dataset is 10 UAV images and 20 satellite images. These images have a spatial resolution of 2 cm and these are taken from distinct cities. When R-CNNs models are evaluated with other existing models it gives a less precision rate of 86.20%. It produces a high recall rate of 74% and an F1 Score of 0.80. This model generates some false detection and missed detection.

VI. CLASSIFICATION OF MEDICAL IMAGES

Qing Li et al., [16] This paper presented the Medical Image Classification with Convolutional Neural Network. This work proposes the convolutional neural network for image patch classification and comprises random neural node dropout. This CNN model used single convolutional layer architecture to lower the number of parameters and this overcomes the over-fitting issues. For a supple test, this work implemented a neural network toolkit including CNN and RBM with performance quickening using Advanced Vector Extensions (AVX). The datasets used in this work are High-resolution Computed Tomography (HRCT) images; it has 113 sets with 2D regions of interest (ROI) annotated indicting the Interstitial Lung Disease (ILD). The classification of image patches is done in five ILD categories: Normal, emphysema, ground glass, fibrosis, and micronodules. This CNN method gives the greatest classification performance on HRCT lung image patchesof ILD patterns.

Kiran Pai et al., [17] used CNN for the classification of 7 different skin lesions. MNIST-HAM10000 dataset consisting of 10015 labeled images with 7 differenttypes of skin lesions is used to train the model. Convolutional layer kernels and an Adam optimizer witha learning rate of 0.001 are used to train the model. A website is created that provides an interface to accept input, feed it to the model and display the output. This model has an accuracy of 78% on test data but more is needed for real data. CNN model for classifying skin lesions has better performance and reduces human effort.

Miyagawa et al., [18] proposed a CNN architecture for classifying Bifurcation regions in Intravascular Optical Coherence Tomography (IVOCT) Images. The images are preprocessed using global threshold, closing, and gradient morphological operations and hough circle transforms. The CNN architecture represents downsampling and upsampling stages with different layers. Stochastic gradient descent algorithms are used during the training phase. To balance training, class weighting was used. This CNN architecture uses Area Under Curve (AUC) as an evaluation metric and has an AUC of 99.7%. Since the dataset used in this method is different it showed better results compared to other techniques for classifying these types of images.

Shuang Chen et al., [19] presented a Convolutional Pyramid Network to classify coronary artery angiograms using kernels of various sizes in multiple convolutions for enhancing features during training. Datasets are gathered from coronary artery angiograms of 14509 patients belonging to 2 classes of 4 types. Thisnetwork involves 3 parts- convolution, object classification, and loss structure. Deep convolution involves extracting single features and giving feature activation as output. Multi-scale convolution generates global fusion features to enhance the richness of the features and avoids information loss. It performs ReLU operation and uses bilinear interpolation to increase nonlinearity and unsample features of small size. Compared to ResNet101, this model improves precisionby 19.1% and 30.5% for 2 class classifications. For multi-class classification, the network convergence rateis faster and training loss decreases at each epoch. This method has a stronger generalization ability.

Juan Lyu et al., [20] present the classification of lung nodules using Multi-Level convolutional neural networks (ML-CNN). It has 3 CNNs to extract the features of lung nodule CT images because these lung nodules have various morphologies and sizes. Then they ground the output of the pooling layer into a one-dimensional vector for every level and join them. This model is applied in three levels namely benign, indeterminate and malignant lung nodules. This work is employed on the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database. It has 2 convolutional layers as Batch normalization and pooling layers. BN is used to reduce internal covariate shifts. The pooling layer is used to produce 60 feature maps in all the levels. Compared to MC-CNN, ML-CNN got the highest accuracy 84.81%.

VII. CONCLUSION

In our current study, AlexNet, VGG, GoogLeNet, PCNN, ResNet, DenseNet, and LAHNet are some of the CNN models used for image classification. These CNN models can handle large datasets without human supervision and extract features by reducing high dimensionality without losing any information. This makes it robust and has high performance in extracting essential features compared to other traditional approaches, thus making it a strong contender for image classification with high prediction accuracy.

References

[1] Zezhong Zheng, Yameng Zhang, Liutong Li, Mingcang Zhu, Yong He, Minqi Lig, Zhengqiang Guoa, Yue Hea, Zhenlu Yu, Xiaocheng Yangh, Xin Liug, Jianhua Luo, Taoli Yang, Yalan Liu, and Jiang Li, "Classification Based On Deep Convolutional Neural Networks With Hyperspectral Image". In Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2017.

[2] Asia Kausar, Mohsin Sharif, JinHyuck Park and Dong Ryeol Shin Sungkyunkwan University, South Korea, "Pure-CNN: A Framework for Fruit Images Classification ". In Proceedings of IEEE International Conference on Computational Science and Computational Intelligence (CSCI) 2018.

[3] Guoqing Ding, Yan Song, Jia Guo, Chen Feng, Guangliang Li, Bo He, the Ocean University of China and Tianhong Yan China Jiliang University, China, "Fish Recognition Using Convolutional Neural Network". In Proceedings of IEEE conference OCEANS 2017 - Anchorage conference September 2017.

[4] Aws Anaz and Marjorie Skubic, University of Missouri, USA, Jay Bridgeman, University of Missouri School of Medicine, MO and David M. Brogan, Washington University School of Medicine, MO, "Classification of Therapeutic Hand Poses Using Convolutional Neural Networks ". In Proceedings of IEEE 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018.

[5] Hui Li, Chinese Academy of Sciences, China, Baoxin Hu, and Qian Li, York University, Canada and Linhai Jing, Chinese Academy of Sciences, China, "Cnn-Based Tree Species Classification Using Airborne Lidar Data And High-Resolution Satellite Image". In Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2020.

[6] Le Kang, Jayant Kumar, Peng Ye, Yi Li and David Doermann, University of Maryland, USA, “Convolutional Neural Networks for Document Image Classification”. In Proceedings of IEEE 22nd International Conference on Pattern Recognition 2014.

[7] Meet P. Shah and Suyash P. Awate, Indian Institute of Technology (IIT) Bombay and Sougata Singha, Yodlee, “Leaf Classification Using Marginalized Shape Context and Shape+Texture Dual-Path Deep Convolutional Neural Network”. In Proceedings of IEEE International Conference on Image Processing (ICIP) 2017.

[8] Pantelis I. Kaplanoglou and Konstantinos Diamantara, Aristotle University of Thessaloniki, Greece, “Margin-Based Sample Filtering For Image Classification Using Convolutional Neural Networks”. In proceedings of 25th IEEE International Conference on Image Processing (ICIP) 2018.

[9] Ping Han, Zetao Chen and Yi Shuan Wang, Civil Aviation University of China, China and Zheng Cheng, Civil Aviation University of China, China, “Polsar Image Classification Based On Optimal Feature And Convolution Neural Network”. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2020.

[10] Yilong Zhang, Yaoxiang Lu, Haixia Wang, Peng Chan and the Ronghua Laing Zhejiang University of Technology, China, “Automatic classification of marine plankton with digital holography using CNN”. In Proceedings of ELSEVIER conference on Optical Laser Technology 2020.

[11] Angang Du, Zhaorui Gu, Zhibin Yu, Haiyong Zheng and Bing Zheng, Ocean University of China, “Plankton Image Classification Using Deep Convolutional Neural Network with Second-order Features”. In Proceedings of IEEE Global Oceans 2020: Singapore – U.S. Gulf Coast.

[12] David J. Attokaren, Ian G. Fernandes, A. Sriram, Y.V. Srinivasa Murthy, and Shashidhar G. Koolagudi National Institute of Technology, India, “Food Classification from Images Using Convolutional Neural Networks”. In Proceedings of IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017.

[13] K. Kogias, I. Andreadis, K. Dalakleidi and K. S. Nikita National Technical University of Athens, “A Two-Level Food Classification System for People with Diabetes Mellitus using Convolutional Neural Networks”. In Proceedings of IEEE 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018.

[14] Yuansheng Hua, Lichao Mou and Xiao Xiang Zhu, University of Munich and German Aerospace Center, Germany, “LAHNET: A Convolutional Neural Network Fusing Low- And High-Level Features For Aerial Scene Classification”. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2018.

[15] Tianyu Tang, Shilin Zhou, Zhipeng Deng, Lin Lei, and Huanxin, Zou National University of Defense Technology, China, “Fast Multi Directional vehicle detection on Aerial images using region-based Convolutional Neural Network”. In Proceedings of IEEE International Geoscience and Remote Sensing Symposium 2017.

[16] Qing Li, Weidong Cai and David Dagan Feng, University of Sydney, Australia, Xiaogang Wang, The Chinese University of Hong Kong, Hong Kong, Yun Zhou, University School of Medicine, United States and Mei Chen, Carnegie Mellon University, United States, “Medical Image Classification with Convolutional Neural Network”. In Proceedings of IEEE 2014.

[17] Kiran Pai IEEE Computational Intelligence Society and Anandi Giridharan Indian Institute of Science, India, “Convolutional Neural Networks for classifying skin lesions”. In Proceedings of IEEE Region 10 International Conference TENCON 2019.

[18] M. Miyagawa, M. G. F. Costa, M. A. Gutierrez, J, P.G.F. Costa and C.F. F. Costa Filho, Member, IEEE, Federal University of Amazonas, Brazil, “Using Convolutional Neural Networks for Classification of Bifurcation Regions in IVOCT Images”. In Proceedings of 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) July 2019.

[19] Shuang Chen, Yang Tang, Xiaotong Shi and Honggang Zhang are with the Beijing University of Posts and Telecommunications, CHINA and Lihua Xie and Bo Xu are with Fu Wai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences, CHINA, “Convolution Pyramid Network: A Classification Network on Coronary Artery Angiogram Images”. In Proceedings of 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) July 2020.

[20] Juan Lyu, Sai Ho Ling, Senior Member, IEEE, “Using Multi- level Convolutional Neural Network for Classification of Lung Nodules on CT images”. In Proceedings of IEEE 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018.