Normalized Gradient Descent based Deep Convolutional Neural Network for Image Classification

Ranjana Kapila
CGC Landran, Mohali

Abstract--The proposed DCNN based image recognition model has been created for the purpose of image classification, which is eventually known as the object recognition in the image processing models. This image recognition model is trained with a larger number of images, where nearly 50 classes of images are present in the given dataset. The proposed model is aimed to achieve the higher recognition accuracy, which is measured in the terms of Word Error Rate (WER), Character Error Rate (CER) and Root Mean Squared Error (RMSE). The proposed model utilizes Deep Learning approach, in which the Convolutional Neural Network (CNN) is deployed with multiple layers to predict the classes in the image data along with the dynamic visual features based upon gradient descent under the neural network to generate the series of linked tensors. In this paper, the proposed DCNN model is tested with different number of layers for pattern recognition. Its variants are designed with 4, 5 and 6 layers. The comparative analysis of the proposed intelligence feature recognition based DCNN 6-layered model is performed in this section, with the existing 6 layer DCNN model. Deep Learning mechanism is employed to recognize the image data. The proposed model has outperformed all of the image recognition contenders for the image classification and object recognition.

Keywords: Image classification, object recognition, Deep neural network, Convolutional neural network.

I. INTRODUCTION

Object recognition and image classification, the subsets of Digital Image Processing, is the intentional alteration of image matrices by using visual features or object shape recognitions. A digital image is a representation of a visual scene in the form of digital data. An image is represented as a continuous variation of instantaneous variation with respect to variations in visual perception, which is described in the form of colours. We know that computers work in Digital domain. So for convenient alteration, storage, retrieval and transmission of an image, it is converted to Digital form. A camera sensor converts the visual perception into the digital form by using the layered deployment of the algorithms.

The general methodology for object recognition is as follows:-

- First the image is pre-filtered using Band-pass or Band-reject filters. Filtering is done to remove the unwanted frequencies (noise) from the image pixels.
- Secondly, the filtered image is cut into a number of smaller segments. A windowing technique having predefined window size (in ms) and hop size/window shift (in ms) is employed for analysing the image block by block from the start till the end.
- After processing the image, a local maxima selection algorithm is employed to select the best pitch candidate. The brightest neuron selected by this algorithm represents the predicted class of the target object in the given test image.

II. LITERATURE REVIEW

Girshick, Ross et. al. (2016) has worked on the region-based feature extraction for the convolutional networks to achieve the higher accuracy of classification. In this paper, the authors have proposed the simple and scalable detection algorithm that improves mean average precision (mAP) by more than 50% relative to the previous best result on VOC 2012—achieving a mAP of 62.4%.

Fleites, Fausto et. al. (2015) has worked on the enhancement of the product detection with multicue optimization for TV shopping applications. The proposed model has been designed to realize this use case and detect the products in the content stream must be detected so that the TV system notifies consumers of possibly interesting ones. **Alireza Masoudian et al. (2013)** said that image classification and segmentation are the two main important parts in the 3D vision system of a harvesting robot. Regarding the first part, the vision system aids in the real time identification of contaminated areas of the farm based on the damage identified using the robot's camera. To solve the problem of identification, a fast and non-destructive method, Support Vector Machine (SVM), is applied to improve the recognition accuracy and efficiency of the robot.

Parisa Pouladzadeh et al. (2012) presented a new recognition algorithm for emerging food classification. The algorithm considering its shape, colour, size, and texture characteristics. Using various combinations of these features, a better classification will be achieved. Based on our simulation results, the proposed algorithm recognizes food categories with an approval recognition rate of 92.6%, in average.

Meenu Dadwal et. al (2012) representing different techniques to detect the rate of ripeness of fruits and vegetables. This paper reports techniques like histogram matching, clustering algorithms based image segmentation and relative value of parameter based segmentation.

B.Sathya Bama et al. (2011) proposes an efficient computer-aided Plant Image Retrieval method based on plant leaf images using Shape, Color and Texture features intended mainly for medical industry, botanical gardening and cosmetic industry. Here, we use HSV colour space to extract the various features of leaves.

III. EXPERIMENTAL DESIGN

The proposed work based upon the image recognition in the aurora dataset has been proposed in this paper. The proposed work has been based upon the appearance based image recognition and classification over the given dataset. The knowledge discovery based image recognition and localization method has been utilized for the image recognition in the proposed work for the word detection or classification.

The Deep Convolution Neural Network (DCNN) classification model has been utilized for the image classification from the given dataset collected from the online source across the internet. The object is detected and classified in the proposed work by using the pattern matching based upon the DCNN under this technique. The latter object displays the image formation of the testing image. The proposed work has been evaluated for its performance under the chain of experiments using this technique. The image recognition method incorporates the knowledge-driven approach, which utilizes the slider window function based pattern discovery within the image matrix to localize the object or strong visual features across the image matrix. The objects are further classified using the second level pattern classification model after the local feature extraction by region localization algorithm under the DCNN model. The DCNN is based upon the feed-forward paradigm in the neural networks in order to compute the likelihood between the training and testing samples. The neural network weight computation for the image recognition has been incorporated under this proposed work to evaluate the probability of the found match in order to determine the final decision. The flexible weight tracking has been applied using the neural network for the purpose of gradient descendent based decision logic calculation:

$$Q_{ij}(t+1) = Q_{ij}(t) + n\frac{\partial b}{\partial q_{ij}}$$

Where the b defines the cost function, Q defines the training coefficient, n is the number of input variables, ∂ denotes the initial probability and t defines the current index count. The supervised methods require the weight calculation to determine the best match out of the available matches from the given set of training vectors. The softmax function is prominently utilized to evaluate the neural weights by using the optimal cost function as it defined in the following equation.

$$q_j = \frac{\exp(y_j)}{\sum_k \exp(y_k)}$$

Where $q_j$denotes the probability of the class and y_k and y_k gives the overall number of unit inputs to the neural network denoted by the variables j and k. The cost entropy function plays the vital role in the successful execution of the neural network classification. The cross entropy mechanism is learned to outperform the Mean Squared Error

(MSE) and classification error, because of its property to inspect the input patterns, and describe the strongest patterns for the likelihood evaluation. The cross entropy based error can be satisfied with the following equation:

$$C = - \sum d_j \log (p_j)$$

Where d_j denotes the final or target probability for each of the unit given by j in the data. The p_j symbols give the final output likeliness or probability for each of the unit given by j in the iterative manner, once the neural network is activated. In this algorithm, the application of knowledge-driven feed forward has been incorporated for the image recognition and classification over the given image dataset.

IV. RESULT ANALYSIS

The figure 1 shows the results of DCNN model with 6 layers for robust object recognition over the 2000 training samples and 125 testing samples, where the neural network is designed in the layered approach with 4000, 2000, 1000, 500, 250 and 125 neurons in the 6-layered model. The average value of RMSE has been recorded at 0.635, which is equal to the DCNN 6-layered model. The average CER has been recorded at 7.87%, whereas the average RMSE is recorded between 0.22 and 0.88 with average value of 0.635. The average WER of 2.8% is recorded over the 10 iterations, and the WER of 2.8% shows the very significant results in case of this model with 6-layers (DCNN 6-layer) with 4000 training samples.

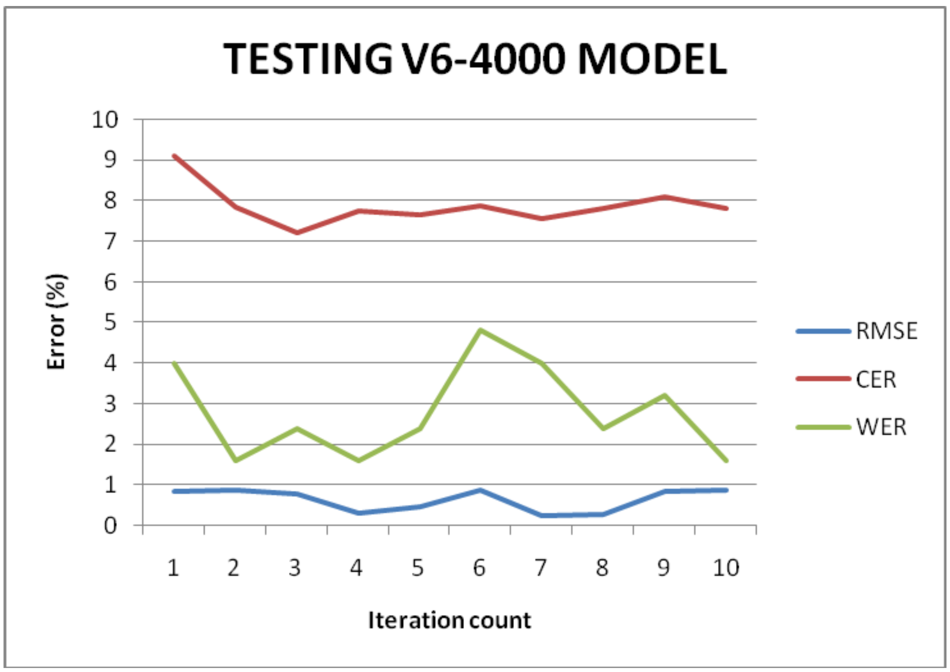

Figure 1: DCNN with 6-layers and 4000 samples

The WER error is recorded between 1 and 5 percent over 4000 sample database and CER between 7 and 9 percent, which shows the higher accuracy in order to recognize the images or objects. The RMSE error also shows the significant results, which is below 1 percent and shows the highly significant results alongside WER and CER.

In the figure 2, results of DCNN with 6-layered model and 8000 training samples are displayed. The average value of RMSE is recorded at 0.576, whereas CER is recorded at 15.694% along with WER at 1.64%. The average value of 1.64% for WER shows the higher significance of this image recognition model. Also the RMSE value shows the similar results and proves the significance of DCNN 6-Layered model for robust image recognition, whereas the CER has been recorded quite higher at nearly 15% error.

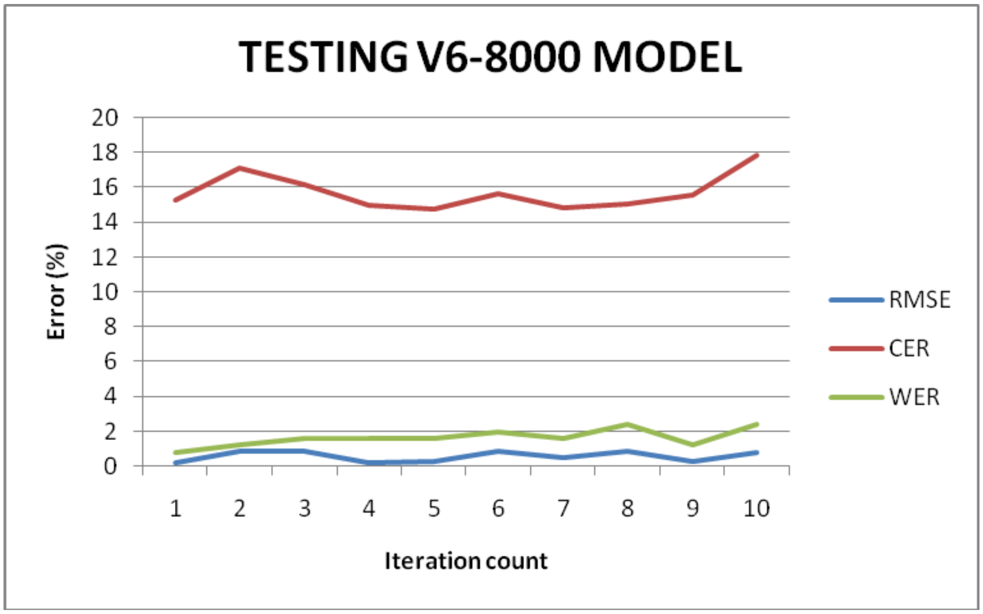

Figure 2: DCNN with 6-layers and 8000 samples

In the following figure 2, the CER is recorded between the 14 and 18, whereas the RMSE and WER are recorded between 0 and 3% in the case with 6-layers DCNN. In figure 3, the performance of different image recognition models is analyzed, which includes the proposed work variants with 4, 5 and 6 layers under the VDL mechanism. The performance analysis survey reveals the higher accuracy of 6-layered DCNN model, which is tested between 2000 and 8000 training samples and 60 to 250 testing samples.

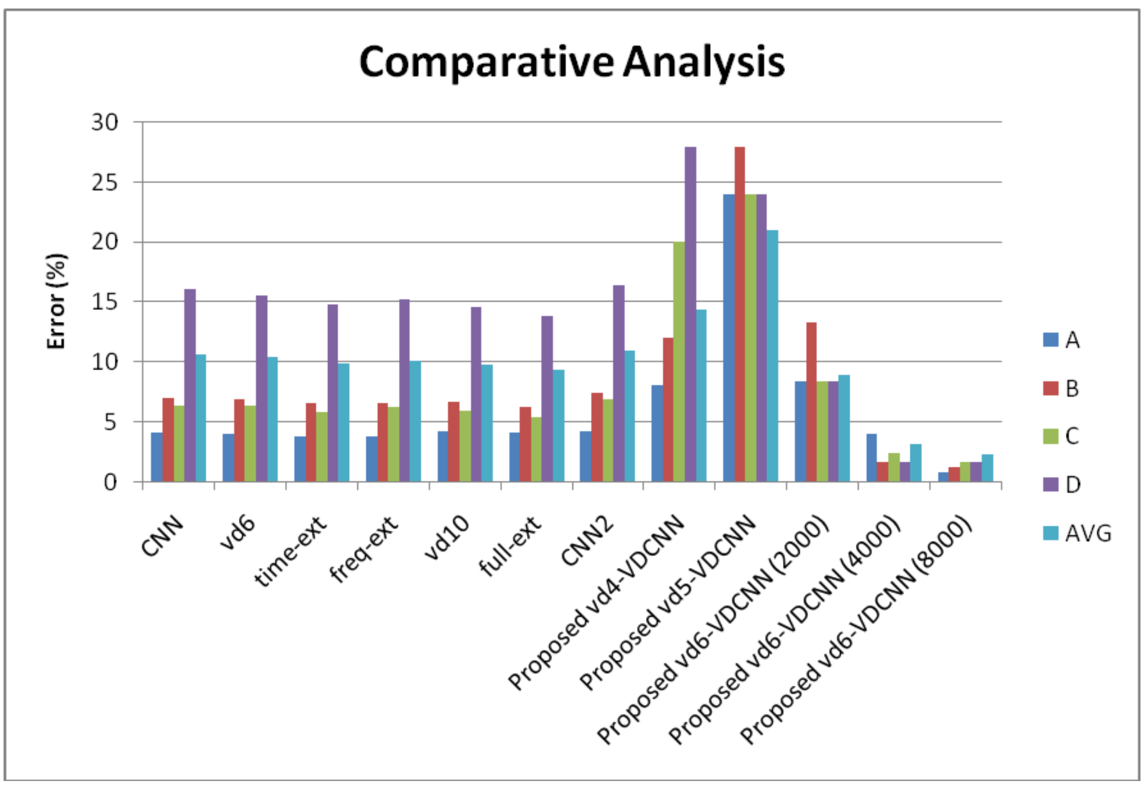

Figure 3 Comparative analysis of proposed model against existing models

In the figure 3, the performance of the proposed model is analyzed among the various models using the randomly drawn testing data. The proposed GMM-MFCC-VDCNN model with 6-layers, is tested with 2000, 4000 and 8000 training samples, and recorded WERs are 7.5%, 2.8% and 1.64% respectively. The average value of WER for all of the GMM-MFCC-VDCNN-6 models is 3.98%, which is highly significant and acceptable. The average WER 3.98% of proposed GMM-MFCC-VDCNN is significantly lower than the baseline models of time extension (time-ext), vd10 and full-ext which have WER of 9.84%, 9.78% and 9.28% respectively. This proves the higher accuracy of the proposed model with different between 2.34% and 7.64%.

V. CONCLUSION

In the case of DCNN 6-layered based image recognition model, the performance analysis is carried out with 2000, 4000 and 8000 samples. The performance analysis of DCNN with 4-layered model concludes the results between 0.2 and 0.33 for RMSE, 1.7 and 2.17 for CER and 8 and 28 for WER, where WER shows the insignificant value but CER and RMSE describes the robust performance on micro levels. The DCNN 5-layered model recorded 0.689 RMSE, 1.78 CER and 20 WER on an average, which shows the higher accuracy on the basis of RMSE and CER and lower accuracy on the basis of WER with 20% error. The DCNN models with 6-layered fashion have added the robustness to the recognition models, where the recognition error of 7.5%, 2.8% and 1.64% is noted with the 2000, 4000 and 8000 samples respectively. The DCNN 6-layered model recorded average WER of 3.98% which has outperformed the existing models CNN2 (10.96%), full extension (9.28%), vd10 (9.78%), frequency extension (10.02%), time extension (9.84%), vd6 (10.34%) and CNN (10.64%). The proposed very deep CNN model has been found improved than the existing CNN models (CNN and CNN2) with 10.64% and 10.96% WER in comparison to the proposed model value of 3.98%.

REFERENCES

[1] Fleites, Fausto C., Haohong Wang, and Shu-Ching Chen. "Enhancing Product Detection With Multicue Optimization for TV Shopping Applications."*Emerging Topics in Computing, IEEE Transactions on* 3, no. 2 (2015): 161-171.

[2] A Plebe, G Grasso, “Localization of spherical fruits for robotic harvesting” Machine Vision and Applications, 2001.

[3] Alfatni, Meftah Salem M., Abdul Rashid Mohamed Shariff, Helmi Zulhaidi Mohd Shafri, Osama M. Ben Saaed, and Omar M. Eshanta, “Oil palm fruit bunch grading system using red, green and blue digital number", Journal of Applied Sciences 8, no. 8: 1444-1452, (2008).

[4] AR Jimenez, R Ceres, JL Pons. “A vision system based on a laser rangefinder applied to robotic fruit harvesting” Machine Vision and Applications, 2000.

[5] Bama, B. Sathya, S. Mohana Valli, S. Raju, and V. Abhai Kumar, “Content based leaf image retrieval (CBLIR) using shape, color and texture features”, Indian Journal of Computer Science and Engineering 2, no. 2: 202-211, (2011).

[6] Bucksch, Alexander, and Stefan Fleck. "Automated detection of branch dimensions in woody skeletons of leafless fruit tree canopies." Proc. SILVILASER : 14-16, (2009).

[7] D Margaritis, S Thrun, “Learning to locate an object in 3d space from a sequence of camera images”, IEEE International Conference on Machine Learning 1998.

[8] Eitz, Mathias, Ronald Richter, Tamy Boubekeur, Kristian Hildebrand, and Marc Alexa, "Sketch-based shape retrieval", ACM Trans. Graph. 31, no. 4: 31, (2012).

[9] Esehaghbeygi, A., M. Ardforoushan, S. A. H. Monajemi, and A. A. Masoumi, "Digital image processing for quality ranking of saffron peach." Int. Agrophysics24, no. 2: 115-120, (2010).

[10] Feng, Guo, Cao Qixin, and Nagata Masateru, "Fruit detachment and classification method for strawberry harvesting robot." International Journal of Advanced Robotic Systems 5, no. 1: 41-48, (2008).

[11] Rabatel, A. Bourely, and F. Sevila, F. Juste, “Robotic Harvesting of citrus: A State of the art and development of French Spanish EUREKA project” Proceedings of the International conference on Harvest and Post harvest Technologies for Fresh Fruits and Vegetables, Guanajuanto, Mexico 1995.

[12] Hassan, Siti Noorul Asiah, Nur Nadiah Syakira Abdul Rahman, and Zaw Zaw, "vision based entomology–how to effectively exploit color and shape".

[13] Ishak, W. I. W., and R. M. Hudzari, "Image based modeling for oil palm fruit maturity prediction." Journal of Food, Agriculture & Environment 8, no. 2: 469-476, (2010).

[14] Ismail, Wan Ishak Wan, and Mohd Hudzari Razali, "Hue optical properties to model oil palm fresh fruit bunches maturity index." In Proceedings of World Multi-Conference on Systemics, Cybernetics and Pattern Recognition, pp. 168-173. 2010.

[15] Jiang, L., B. Zhu, X. Cheng, Y. Luo, and Y. Tao, "3 D Surface Reconstruction and Analysis in Automated Apple Stem-End/Calyx Identification." Transactions of the ASAE 52, no. 5: 1775-1784, (2009).

[16] Masoudian, Alireza, and Kenneth A. McIsaac. "Application of support vector machine to detect microbial spoilage of mushrooms." In *Computer and Robot Vision (CRV), 2013 International Conference on*, pp. 281-287. IEEE, 2013.

[17] Pouladzadeh, Parisa, Gregorio Villalobos, Rana Almaghrabi, and Shervin Shirmohammadi. "A Novel SVM Based Food Recognition Method for Calorie Measurement Applications." In *ICME Workshops*, pp. 495-498. 2012.

[18] Dadwal, Meenu, and Vijay Kumar Banga. "Estimate ripeness level of fruits using rgb color space and fuzzy logic technique." *International Journal of Engineering and Advanced Technology* 2, no. 1 (2012): 225-229.

[19] Bama, B. Sathya, S. Mohana Valli, S. Raju, and V. Abhai Kumar. "Content based leaf image retrieval (CBLIR) using shape, color and texture features."*Indian Journal of Computer Science and Engineering* 2, no. 2 (2011): 202-211.

[20] Li, Peilin, Sang-Heon Lee, and Hung-Yao Hsu. "Use of a cold mirror system for citrus fruit identification." In *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on*, vol. 2, pp. 376-381. IEEE, 2011.

[21] Venkatraman, Divya, and Anamitra Makur. "A compressive sensing approach to object-based surveillance video coding." In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 3513-3516. IEEE, 2009.

[22] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Region-based convolutional networks for accurate object detection and segmentation." *IEEE transactions on pattern analysis and machine intelligence* 38, no. 1 (2016): 142-158.