Multiclass Image Classification using CNN and CLAHE Hybrid Model at Runtime

Siddhant Dani, Prof. V. S. Rasal, Hrishikesh Panse, Kshitij Chaudhari, Shruti Kotwal
Computer Engineering Department,
NBN Sinhgad School of Engineering, Pune-411041, India.

Abstract: As data production has expanded at higher pace, the widespread use of automation and surveillance cameras and the requirement for visual feedback on artificially smart gadgets throughout the world has grown significantly the amount of image information that is now created. To simplify the image related jobs, the necessity for efficient image processing has increased. The classification of the categorical image takes thousands of pictures to train the algorithm. The system also needs a long time to extract the characteristics and to classify them. Thus, deep convolutional neural networks were created for image identification tasks and in recent years have shown promising results. CLAHE is a type of adaptive histogram equalization where contrast amplification is limited so that the noise amplification problem is minimised. This paper offers a hybrid CLAHE-deep convolutional neural network architecture for multiclass image classification at runtime. In order to effectively train the model, the suggested architecture is examined using multiple datasets. The experiment employs the Contrast Limited Adaptive Histogram Equalization (CLAHE) architecture as a pre-processing step for picture enhancement, followed by the CNN architecture for image classification. In terms of accuracy obtained, the experimental results imply that the suggested hybrid design beats traditional techniques.

IndexTerms - Image Classification, CNN, Deep Learning, Image Processing, Deep Networks, Neural Networks, CLAHE, Contrast Limited Adaptive Histogram Equalization.

I. INTRODUCTION

Image classification is an essential problem in artificial computer vision that has sparked a lot of attention in recent decades. This field attempts to categorise an input picture based on its visual content. To characterise a picture in a discriminatory manner, most researchers now rely on hand-crafted features, HoG or SIFT. Then, to reach a final choice, learnable classifiers such as SVM, random forest, and decision tree are applied to the retrieved characteristics. However, when a large number of photos are provided, it becomes a challenging task to extract characteristics from them.

This is the one of reasons that deep neural network model is coming into the picture. Deep learning algorithms are neural network-inspired systems that leverage enormous data sets to anticipate outcomes in a semi-supervised learning context. Image categorization accuracy with CNN outperforms any machine learning approach and even other deep learning methods. The notion of deep learning was originally described in 2006. And it hasn't stopped evolving since then. Deep learning is defined as a class of machine learning algorithms that use numerous layers of non-linear information processing for supervised or unsupervised feature extraction and modification, as well as pattern analysis and classification.

Sometimes, the collected photos are hampered by various flaws caused by a variety of unpredictable situations such as interior illumination, evening lighting, and gloomy weather. Because of the existence of noise, the tangent surface's reflection is poor under these conditions, and colour is obliterated. To overcome this problem, we use an algorithm called CLAHE. Adaptive histogram equalisation with contrast limitations is a modified version of adaptive histogram equalisation. The enhancement function is applied to all neighbouring pixels in this technique, and a transformation function is created. Because of the contrast limiting, this differs from AHE.

The paper is structured as follows: Section II gives information on the state-of-art work done in past related to Image Classification. Section III presents the proposed approach with detailed description. Later, the experimental results are presented in section IV followed by the conclusion in section V.

II. OVERVIEW AND RELATED WORK

There have been multiple approaches discussed in the past to perform image classification tasks. In literature there exists many papers that have proposed different techniques with decent accuracy. Many papers focused on theoretical as well as practical aspects of image processing in the field of image classification.

In [1], a CLAHE and CNN based model was described to improve the early diagnosis process for breast cancer. The proposed model was experimented with the BreakHis dataset containing breast microscopic images. The experiment is carried out using the Contrast Limited Adaptive Histogram Equalization (CLAHE) hybrid architecture, followed by the watershed method, as an improvement for image pre-processing, followed by the image classification CNN architecture. The CNN architecture proposed was LeNet-5 which comprises of total of 8 layers. They performed a comparison between the normal CNN based system and the upgraded hybrid system. And they were able to conclude that the hybrid system is much better than the traditional CNN based system. They were able to achieve an accuracy of 90% which is greater than the accuracy of traditional approach by 3%.

In [2], the authors proposed the two models comprising of CNN and SVM each. The CNN architecture used was LeNet5. They created their own custom datasets for training both the models which contained over 3000 images in CNN dataset and 350 images for SVM dataset. They used python NumPy library to flatten all the images into an array of numbers which made it much easier to use the images for training and testing purposes. Using SVM their model was able to achieve an accuracy of 82% while the CNN model was able to reach the accuracy of 93.57%. By conducting this experiment, they were able to conclude the superiority of CNN over traditional approaches.

In [3], the authors were trying to determine physical features of vehicles using image processing and deep learning. They used the data obtained from traffic surveillance cameras in order to train the model. The model defined used grey conversion and background subtraction as pre-processing steps while Deep Neural Classifier for classification. They built the model in TensorFlow and trained it in the offline mode for thousands of images. They were able to classify images into 7 categories with more than 90% accuracy for each class. Similarly, another approach was described in [4], which uses Capsnet to classify images from MNIST and CIFAR10 datasets. Capsnet is similar to a CNN in terms of architectural design. Upon using this model on MNIST dataset they were able to achieve an accuracy of 99.71% and on CIFAR10 dataset they achieved an accuracy of 73.3%. But after conducting the study they were able to conclude that after a few modifications to initial CNN architecture the newly obtained Capsnet can perform better than CNN.

In [5], the authors used CLAHE algorithm to enhance the quality of video footage obtained in Real time. They proposed this enhancement to tackle the problem of foggy footage or hazy image. This helped them improve the performance of Real Time Surveillance systems by a great factor. They were able to reduce the time required for processing of single frame from 8.6 seconds to 6.8 seconds.

In [6], they proposed a model based on RESNET and SVM, which used RESNET for feature extraction and based on those extracted features used SVM to classify the images into respective categories. They described two types of RESNET's which have 18 and 34 layers respectively. The proposed model with 18 layers and SVM classifier was able to achieve 93.57% accuracy and the model with 34 layers and SVM classifier was able to achieve 91.58% accuracy. At the end the authors were able to conclude that the newly designed RESNET has much better performance than a CNN. But they also found out that adding more layers to RESNET reduces its performance.

In [7], the model described was a small 13-layer simple convolutional neural network. The dataset on which it was trained contained 100,000 images from 200 classes. They also introduced a ReLU function to bring non-linearity to the model, thus improving its performance. After 100 iterations the model was able to return 99% training and 96% validation accuracy. But since the model is applied only once to the dataset the training process was carried out for a long time. In [8], the proposed system uses HOG+SVM and Alexnet transfer learning algorithm. Alexnet uses transfer learning to train the model. Here HOG is used as a feature extraction method. The HOG features obtained are supplied to the SVM for obtaining pre-classification results and then Alexnet is used to classify images into different classes. The system used was able to obtain an accuracy of 95% and also was able to avoid the problem of overfitting in deep learning.

In [9], the proposed model tried to solve the classic cat-dog classification problem. They used a 3 layered CNN for classification as well as feature extraction. In the proposed architecture the CNN for feature extraction consisted of alternate convolutional and pooling layers. At the end of 50 epochs the model was able to reach an accuracy of 84.45% and also solved the issue of overfitting due to large dataset. They also concluded that the performance can be improved more by adding more layers to the CNN. In [10], the authors used the V3 inception model of CNN to solve the most challenging problem in image classification i.e., Food Classification. They used the food-11 dataset to train the model. The V3 inception model used consists of 7 layers in the CNN architecture. The proposed model after a few epochs achieved an accuracy of 92.86%. But at initial iteration overfitting was observed. The authors also concluded that by using feature-based models the computation time can be reduced significantly. Also, some other algorithms like RNN and DCNN will give out the similar performance for the given dataset.

The primary concerns of the writers regarding the dataset were found in the literature study. Data sets from some renowned sources have been discovered in majority of the investigations and which are openly available.

III. Proposed Approach

A hybrid technique is presented to categorise the pictures at runtime, which employs Contrast Limited Adaptive Histogram Equalization (CLAHE) as a pre-processing stage of image enhancement, followed by the VGG-19 deep network. The VGG-19 is a variation of Convolutional Neural Network (CNN) that is used to classify pictures. The basic system architecture is given in the figure 1 below.

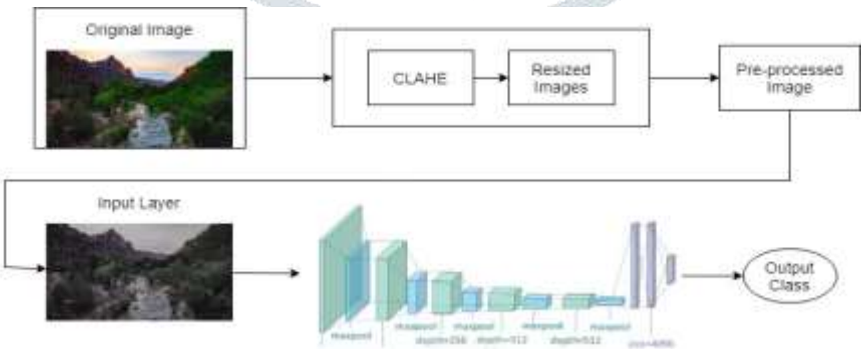

Fig. 1: Basic System Architecture

III.A CLAHE

In order to minimise the problem of noise amplification, CLAHE which is a type of histogram adaptive equalization where the contrast amplification is minimised is employed. Histogram equalisation is a contrast adjustment image processing technique employing the pixel-value histogram. It boosts the global picture contrast, particularly when the picture has close contrast values. It is used to improve the details of a picture. Equalization of histogram when the histogram of the picture is constrained to a certain region is regarded effective. This might not function well if huge intensity swings occur and histograms span a wide area, i.e., both dark and bright pixel areas. To resolve this, AHE is used. The whole image is broken into small, single blocks called "tiles" in this process. Then in each of those tiles histogram equalization is performed.

- Steps for CLAHE Algorithm:

1. The first step is to acquire an input picture.

2. In second step, obtain all input values needed in the improvement process, such as the number of regions in the row and column directions individually, the dynamic range (number of bins utilised in the histogram transform function), cliplimit, and distribution parameter type.
3. In third step, Divide the picture into 2x2 contextual sections i.e., Tiles
4. In fourth step, use local histogram equalisation on each tile (apply SHE to each tile).
5. In fifth step, set a threshold limit to prevent the image from being oversaturated (in homogenous regions). These regions are distinguished by a high peak in the histogram as a result of a large number of pixels falling within the same grey level range.
6. In sixth step, create a clipped histogram and grey level mapping for each tile.
7. In next step, modification of the histogram is done by computing the slope of the cumulative distribution function produced by adding the counts of histogram. A bigger slope correlates to a higher bin count. Thresholding guarantees that the maximum permitted pixel count is limited in such a manner that the height of each bin does not exceed the clip limit.
8. In eighth step, in order to generate a better picture, interpolate grey level mapping. In this procedure, four-pixel clusters are used and a mapping process is conducted. Each of the mapping tiles will partially overlap in the picture region, after which a single pixel will be removed and four mapping will be performed to that pixel.
9. In the last step, interpolate between those findings to obtain an enhanced pixel, then repeat across a picture.

III.B Deep Learning Model, VGG-19

A convolutional neural network (CNN) is a type of neural network that extracts or recognises a certain feature in an image. This is one of the most fundamental processes in Machine Learning, and it's commonly used as a foundation model in most Neural Networks, including GoogleNet, VGG19, and others, for tasks like object detection, image classification, and so on.

The VGG19 model is a variation of the VGG model that has 19 layers in total (16 convolution layers, 3 Fully connected layer, 5 MaxPool layers and 1 SoftMax layer). VGG19 has a FLOP count of 19.6 billion.

This model takes a fixed size RGB image i.e., 224*224. Hence the shape of the matrix in this network is 224*224*3. The only pre-processing this model performs is that it subtracts the mean RGB values from each pixel which computed over the entire dataset. The kernels used are of size 3*3 with stride equal to 1, which enables the model to cover the entire picture notion. It uses spatial padding to maintain the original resolution of the image. The model applies the maxpool layer on a 2*2-pixel window with stride equal to 2. Then a Rectified linear unit (ReLU) function was introduced in order to bring non-linearity to make the model classify better and to improve the computational time taken by the model. As compared to other CNN models introduction of ReLU proved that the model training time can be reduced significantly. It consists of three fully connected layers of size 4096 and after these layers it has one final softmax layer which has 1000 channels for 1000-way image classification.

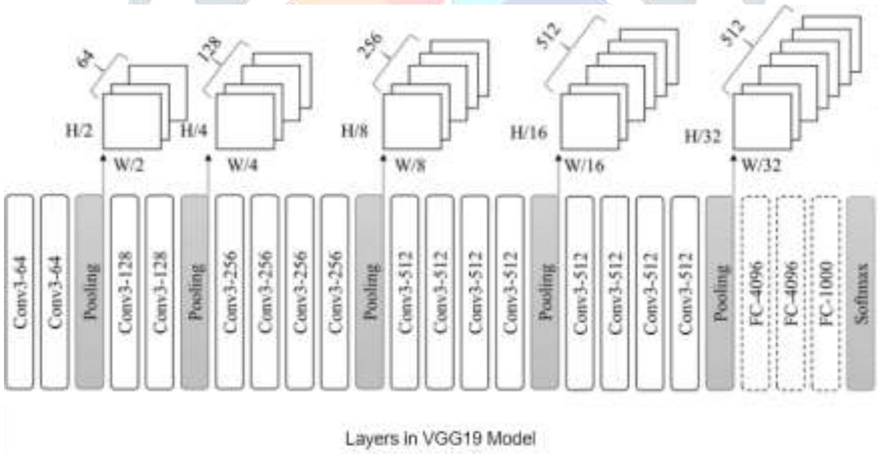

Fig. 2: Original VGG19 Architecture

IV. Experimental Results

IV.A Dataset description

The dataset is chosen from a freely available source, Kaggle and University of Toronto Repository that is deemed a state-of-the-art data repository for the experimentation of the proposed methodology. The datasets chosen were Fashion-MNIST and CIFAR-10. Fashion-MNIST is a dataset of Zalando's article photos, including 60,000 instances in the training set and 10,000 examples in the test set. Similarly, CIFAR-10 is another dataset for classifying images. It comprises of 60,000 instances belonging to 10 classes of pictures. A total of 50,000 pictures and 10,000 test photos are available. The dataset consists of six sections - five training sets and a single test lot. There are ten thousand pictures in each batch. Each sample from both the datasets is a 28x28 grayscale picture with a label from one of ten classes.

Also, we used another custom-made dataset to test the working of our proposed methodology which consists of images of random sizes obtained from sources like Google, phone cameras and end-users. It has more than 10000 images belonging to different classes.

IV.B Results

In order to test our model, we used the training and testing batches from the datasets obtained from the web. We split Fashion-MNIST dataset into 48000 training samples and 12000 validation samples and 10000 testing samples. Also, we split the CIFAR-10 dataset into 35000 training samples, 15000 validation samples and 10000 testing samples. But for our custom dataset we split the dataset into two parts, training set and testing set. We split the dataset using 80-20 split ratio. The experimentation results are shown below in figure.

VGG19 with CLAHE

Dataset	Training Accuracy	Testing Accuracy	Validation Accuracy
FASHION-MNIST	0.7010	0.7664	0.7605
CIFAR-10	0.9645	0.8508	0.8425
CUSTOM	0.9552	0.7908	0.7851

Fig. 3: Performance Results

Fig. 4: Original Image

Fig. 5: Image with CLAHE Pre-processing

Fig. 6: Original Image

Fig. 7: Image with CLAHE Pre-processing

V. Conclusion

In this research paper, a new hybrid technique for picture classification at runtime is suggested that is based on CLAHE with VGG19 deep neural network architecture. The datasets used are collected from freely available sources such as Kaggle, Google and Uni. of Toronto Repository, named Fashion-MNIST and CIFAR-10 having 10 classes and 60000 instances. The proposed system achieves the accuracy of 76% for classifying the Fashion-MNIST dataset, 85% on CIFAR-10 dataset and 78% on the custom created dataset. The suggested technique employs pre-processing stage in the form of CLAHE, which improves the outcome even further. Having said that, there is still opportunity for development, not just in terms of accuracy, but also in terms of additional sub-classification of the pictures. Image categorization and deep neural networks are both very fresh domains that are still being investigated.

References

[1] Ankit Vidyarthi, Jatin Shad, Shubham Sharma, Paridhi Agarwal, "*Classification of Breast Microscopic Imaging using Hybrid CLAHE-CNN Deep Architecture*," in 2019 IEEE.

[2] Sai Yeshwanth Chaganti, Ipseeta Nanda, Koteswara Rao Pandi, Niraj Kumar, *"Image Classification using SVM and CNN*," in 2020 IEEE Xplore.

[3] Turan Goktug Altundogan, Mehmet Karakose, "*Image Processing and Deep Neural Image Classification Based Physical Feature Determiner for Traffic Stakeholders*," in 2019 IEEE.

[4] Zhiyong Dong, Sheng Lin, "*Research on image classification based on Capsnet*," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, IEEE, 2019.
[5] Garima Yadav, Saurabh Maheshwari, Anjali Agarwal, "*Contrast Limited Adaptive Histogram Equalization Based Enhancement for Real Time Video System*," IEEE, 2014.
[6] Mrs. Arpana Mahajan, Dr. Sanjay Chaudhary, "*Categorical Image Classification Based on Representational Deep Network (RESNET)*," in 2019 IEEE Third International Conference on Electronics Communication and Aerospace Technology, IEEE, 2019.
[7] Shyava Tripathi, Rishi Kumar, "*Image Classification using small Convolutional Neural Network*" in 2019, IEEE.
[8] Yunyan Wang, Chongyang Wang, Lengkun Luo, Zhigang Zhou, "*Image Classification Based on transfer Learning of Convolutional neural network*" in 2019 38th Chinese Control Conference, IEEE, 2019.
[9] Santisudha Panigrahi, Anuja Nanda, Tripti Swarnkar, *"Deep Learning approach for Image Classification,"* in 2018 2nd International Conference on Data Science and Business Analytics, IEEE, 2018.
[10] Md Tohidul Islam, B.M. Nafiz Karim Siddique, Sagidur Rahman, Taskeed Jabid, *"Image Recognition with Deep Learning,"* in 2018 ICIIBMS, IEEE, 2018.
[11] Siddhant Dani, Prof. P. S. Hanwate, Hrishikesh Panse, Kshitij Chaudhari, Shruti Kotwal, *"Survey on the use of CNN and Deep Learning in Image Classification,"* in JETIR, 2021.
[12] Manju R. A, Koshy G, Philomina Simon, *"Improved Method for Enhancing Dark Images based on CLAHE and Morphological Reconstruction,"* in ScienceDirect, 2019.
[13] Brij Bhan Singh, Shailendra Patel, *"Efficient Medical Image Enhancement using CLAHE Enhancement and Wavelet Fusion,"* in IJCA, 2017.
[14] C. Daniel Nesa Kumar, R. Aruna, *"Contrast Limited Adaptive Histogram Equalization (Clahe) Based Color Contrast and Fusion for Enhancement of Underwater Images,"* in IOSRJEN, 2018.
[15] Evgin Goceri, "*Analysis of Deep Networks with Residual Blocks and Different Activation Functions: Classification of Skin Diseases*," in 2019, IEEE.
[16] S. Regina Lourdhu Suganthi, Dr. Hanumanthappa, Dr. S. Kavitha, "*Event Image Classification using Deep Learning*," in 2018 International Conference on Soft-computing and Network Security, IEEE, 2018.
[17] Rupal Agarwal, *"State Classification of Cooking Images Using VGG19 Network,"* University of South Florida, 2019.