Image Classification using SVM and CNN

Sai Yeshwanth Chaganti
Test Engineer
Irdeto Technologies Pvt Ltd
New Delhi, India
yesh.chaganti@gmail.com

Ipseeta Nanda
Electronics and Communication
Koneru Lakshmaiah Education Foundation
Vaddeswaram, India
ipseeta.nanda@gmail.com

Koteswara Rao Pandi
Electronics and Communication
NIIT University
Neemrana, India
pandik.rao@st.niituniversity.in

Tavva GNRSN Prudhvith
Computer Science and Engineering
NIIT University
Neemrana, India
tavvag.narasimha@st.niituniversity.in

Niraj Kumar
Test Engineer
Irdeto Technologies Pvt Ltd
New Delhi, India
nirajkumar513@gmail.com

Abstract —On the surface, teaching a computer to do something like image classification seemed very intriguing to us. Moreover, there are countless real-world applications of this concept. It is in light of these reasons that we decided to work on Image Classification. Thankfully though, this topic has been well-researched by the scientific community and we didn't break a sweat finding resources to learn from. So naturally, we perused a bunch of research papers that dealt with image classification, each from a different perspective. We then decided to implement image classification on a small-scale with the limited hardware we were in possession of. As difficult as it was, we started with SVM and a very small dataset to achieve an accuracy of 93%. Although SVM is a very strong technique, achieving such a high accuracy is still an anomaly. We realized that our results boasted such a high accuracy due to the lack of a large enough dataset. So, using data augmentation, we more than tripled the size of our dataset. On performing SVM now, we achieved an accuracy of 82%, a significant decrease. Unsatisfied with the results, we decided to move to other deep learning techniques. This quest led us to Neural Networks and, CNN. On successfully implementing CNN, we achieved an accuracy of a staggering 93.57% on the very same dataset. This stands as a testimony to the increased potential of deep learning techniques over the more traditional machine learning techniques.

Keywords: Python, Support Vector Machines (SVM) and Convolutional Neural Networks (CNN)

I. INTRODUCTION

To develop an understanding of image classification from the ground-up, i.e., start with the traditional machine learning techniques first and then move onto deep learning techniques. To research the underlying principles and techniques of image classification and to highlight the potential of deep learning techniques and illustrate their importance over the more traditional machine learning techniques. As mentioned before, we began with the more traditional algorithms for classification. After having gone through most of them, we chose to implement SVM. Although it made more sense to implement Neural Networks (NN), the more modern approach, we felt that SVM would be a better starting point and we weren't wrong. Now before we begin with what SVM is and how we implemented it, we think we require to explain why we thought that NN should've been the way to go. Accuracy of Image classification using NN far exceeds any machine learning algorithm and even other deep learning algorithms. The performance of traditional machine learning algorithms has a saturation point and feeding more data beyond this point is not going to result in increased performance. Now that we've expressed the importance of deep learning algorithms, we can move on to SVM. When we began understanding this technique, we were, and this might not come off as a surprise, overwhelmed. Note here that SVM is a very strong classification method by its own merit. But as mentioned before, it does not stand a chance in a data rich world. On further research, dividing the topic into three sub-topics helped us understand more efficiently. The sub-topics are Visualization, Tuning Parameters and Implementation and Analysis. Given a labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which refers to the line after transformation/kernelling. Although we have made a classifier, it is not the most efficient. Using something called the biggest gap trick optimization, SVM finds out the most efficient classifier and support vectors are the major contributors to the calculation of the most efficient classifier, hence other points do not help. After all the above steps, the accuracy is not what we expected, then after some research we thought of Deep Learning where they learn by adjusting the strengths of their connections to better convey input signals through multiple layers to neurons associated with the right general concepts. When data us fed into a network, each artificial neuron that fires transmits signals to certain neurons in the next layer, which are likely to fire if multiple signals are received. The entire process filters out noise and retains only the most relevant features. Moreover, the features given to an NN model are just the pixel values of the image. So, it creates a layer of abstraction for people without a lot of knowledge on image processing. More importantly, its ability to learn features on its own puts it in a league of its own. The remaining paper is organized as follows. Section II introduces Related work. Section III presents our proposed Solution. Section IV explains the Performance Analysis. Results are in Section V. Conclusion and Future work in Section V.

II. RELATED WORK

In [1], this paper attempts to study and provide a brief knowledge of the different image classification approaches and different classification method and discussed different

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Authorized licensed use limited to: University of Southern Queensland. Downloaded on July 04,2020 at 16:48:17 UTC from IEEE Xplore. Restrictions apply.

approaches based on the different types of inputs. In [2], the author put forward image classification as inevitable for pattern recognition, besides that it is the last stage of pattern matching. In addition to this, the classification process delineates the proportion of accuracy in pattern recognition. In [3], the author delineates classification is indeed vital and challenging part in computer science domain. Classification is based on description, similarity of objects, etc. In a nutshell, the pixel is the main block for classification and will be grouped in different classes. In [4], In this paper, CNN is the state-of-art technique for the classification of traffic signs. In addition to this, the results after applying CNN are to learn features and classify RGB-D images task. We use the transfer learning technique namely "Fine-tuning Technique", it mainly depends on reusing layers on trained data to classify our testing set. In [5], the Author states that in this contemporary epoch, the trend for image classification is escalating at an unprecedented rate with various fields such as image processing, computer vision, and machine learning fields. In this paper, we focus on deep learning techniques to solve the image classification problem. In [6], 1.2 million high resolution images are classified by trained CNN into 1000 different classes for ImageNet benchmark. With 650000 neurons the neural network consists of five convolutional layers consist of max-pooling layers, and three fully connected layers. In [7], the ANN and SVM are the two areas applied in image classification. Responsive class by an ANN is result of classification of sub-image. The experimental result shows recognition application and the precision rate as 86%. Which call this model ANN_SVM as this model bring many ANN and one SVM into this proposal. In [8], The previous papers discussed the basic approaches to image classification. Going through them gave us enough insight to move onto more advanced techniques and image classification techniques explored in this survey paper are categorized into supervised and unsupervised learning in which Artificial Neural Networks, Support Vector Machines, Directed Acyclic Graph - SVM, belong to supervised techniques, whereas Fuzzy Decision Trees belong to unsupervised learning. Here, we discuss about three classification methods namely Supervised learning in which the learner is required to learn the behavior of a function by referring to several input-output examples of the function, Unsupervised Learning, labelled examples are not available for reference to learn a function and lastly, Semi-supervised Learning where the training data contains combination of labelled and unlabeled data.

III. PROPOSED SOLUTION

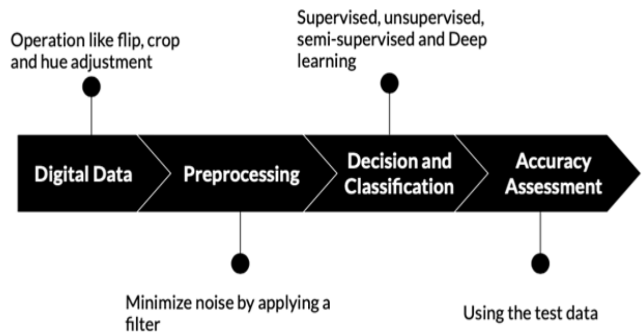


Figure 1: Flow Chart of the Scenario

DIGITAL DATA: At the start of the project a dataset of 350+ images were collected, and all the images were intact with very little noise. For the first implementation that is using the SVM for Image classification, as it is a machine learning technique this dataset was enough to see decent results. We acquired an accuracy of 93%. But these results were not observed when images were a little blurred or have a brighter background. So, it was decided that we would use CNN for the same. The major problem with the dataset if we has to implement CNN is, it requires a lot of data for training the model. So, we explored some of the data augmentation techniques to increase the number of images in the dataset. A few more images were added as well. Making a total of 3000+ images. The different data augmentation techniques we used are Changing the colour space to CMY color space from RGB color space, dithering of the image, Creating Primary Colours Version of the image, Random translation of the image, Flipping the image Vertically, Flipping the image horizontally.

PREPROCESSING:
At the start of the project a dataset of 350+ images were collected, and all the images were intact with very little noise. Every image we get from the dataset goes through the different processes mentioned in the previous section to form more than 7-8 images with all the operations mentioned. This makes a total of more than 3000+ images in the dataset. This dataset is then loaded into the python using a user defined function which first converts all the images to 64 by 64 and then flattens all the images using the NumPy library functions resulting in an array of the numbers for each image. These images or the array of arrays are then used for processing and classification. This type of data is very useful because, it is digital images converted to numbers and these can be easily processed be it be for image classification using SVM, CNN or any other technique.

DECISION AND CLASSIFICATION:
As mentioned before, we began with the more traditional machine learning techniques for classification. After having gone through most of them, we chose to implement SVM. Although it made more sense to implement Neural Networks (NN), the more modern approach, we felt that SVM would be a better starting point and we weren't wrong. Now before we begin with what SVM is and how we implemented it, we think we require to explain why we thought that NN should've been the way to go. Accuracy of Image classification using NN far exceeds any machine

Authorized licensed use limited to: University of Southern Queensland. Downloaded on July 04,2020 at 16:48:17 UTC from IEEE Xplore. Restrictions apply.

learning algorithm and even other deep learning algorithms. More importantly, its ability to learn features on its own puts it in a league of its own. The above figure illustrates exactly this. The performance of traditional machine learning algorithms has a saturation point and feeding more data beyond this point is not going to result in increased performance. This is where NNs shine. Moreover, the features given to an NN model are just the pixel values of the image. So it creates a layer of abstraction for people without a lot of knowledge on image processing. Now that we've expressed the importance of deep learning algorithms, we can move on to SVM.

IV. PERFORMANCE ANALYSIS

After understanding all the tuning parameters, we were pretty much set to implement SVM. We have 2 different implementations that each use a different dataset. The first implementation was using IRIS Dataset to understand how SVM is implemented. Once we understood how, we implemented SVM on our aforementioned dataset, i.e., SVM using Image Pixel Values as Features. Note here that when we implemented SVM, our dataset contains RGB images of 5 different classes. Each image is stored in format [class name] [serial no].jpg'. These are the divisions of 67 images containing dalmatian dog at different ages, animated non animated, with and without background, 53 dollar bill images are cropped to fit the bill, 53 images of pizza slices cropped from advertising posters, with and without background, 65 soccer ball images, 85 sunflower images taken in natural setting, some with the entire plant. The dataset contains a total of 323 pictures after implementing the above techniques and classes are labelled as [0-4]. Now one might ask why we chose SVM over simple clustering. It's because dimensionality reduction using principal component analysis was performed on our feature vector so that it can be visualized graphically. This helped us to understand there are too many overlaps among the classes and hence cannot be clustered. The following PCA scatter plot illustrates the same.

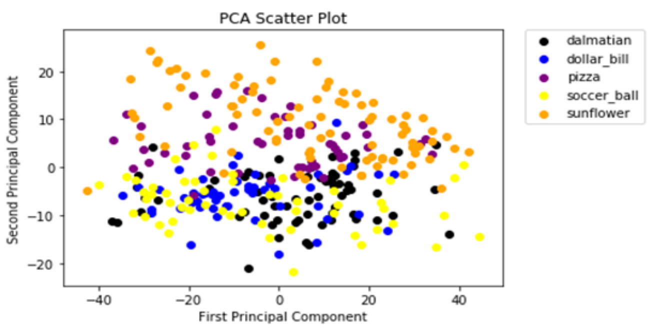


Figure 2: Graph illustrating Disadvantages of clustering.

During implementation, we used two kernels. One was linear kernel as the data was not linearly separable but the accuracy in that case was very low, so we had to shift to RBF kernel which, unsurprisingly resulted in a better F1 score. After conducting experiments on C values {1,10,100,1000} and gamma values {0.001, 0.0001}, we found out that C = 1 and gamma = 0.001 yielded the best results. Following are the results of the implementation:

Classes	Precision	Recall	F1 Score
Dalmatian	0.78	0.82	0.8
Dollar Ball	0.88	0.93	0.9
Pizza	1	0.88	0.93
Soccer Ball	0.88	0.7	0.78
Sunflower	0.83	1	0.91
Micro-Average	0.86	0.86	0.86

Figure 3: SVM Results (Accuracy Assessment)

METHODOLOGY: The following steps were followed for classifying images using CNN. After having successfully implemented SVM, we moved on to Neural Networks, Convolutional Neural Networks (CNN). The CNN Architecture we used is as follows:

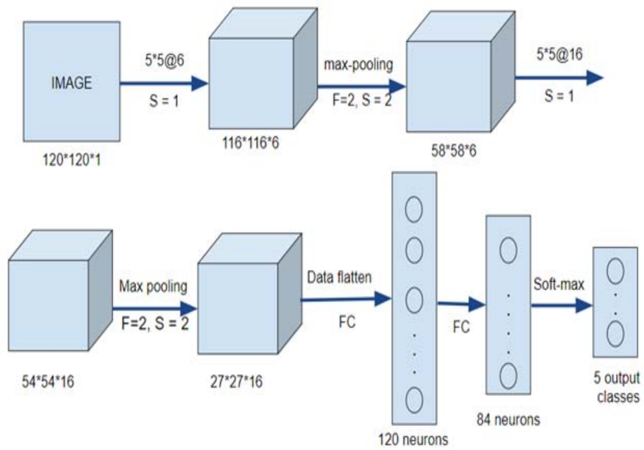


Figure 4: CNN Architecture

- In the proposed methodology, we have used LeNet5 architecture similar to the structure of CNN.
- In this first an image of shape 120x120x1 is convolved with a filter to produce 116x116x6 output.
- On this max-pooling with a stride and size of 2 operation is applied to reduce the output to 58x58x6.
- To which filter and max-pooling of stride and size 2 is applied to result to further reduce the output to 27x27x16.
- This result is then flattened until we obtain a fully connected 84 neurons length layer.
- At last, soft-max is applied which in turn gives us the we the result in terms of the 5 classes of images we have, that is the probability to which class it belongs to.

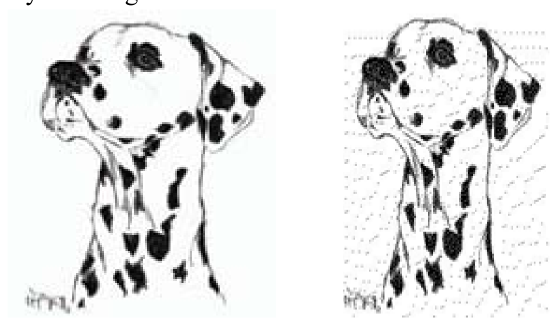
Authorized licensed use limited to: University of Southern Queensland. Downloaded on July 04,2020 at 16:48:17 UTC from IEEE Xplore. Restrictions apply.

V. RESULTS OF THE WORK DONE

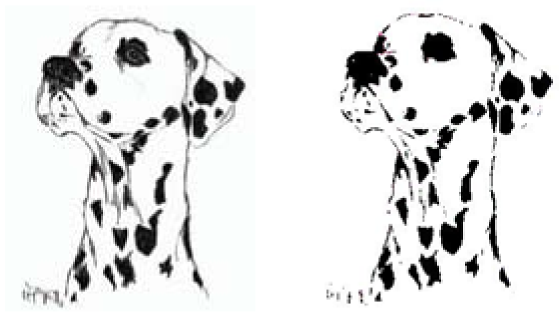
The different data augmentation techniques used are;

Conversion to CMY Color space:
The images in the dataset are in RGB color space and all the 350+ images are converted to CMY color space. It was done using the user defined function which is shown below.

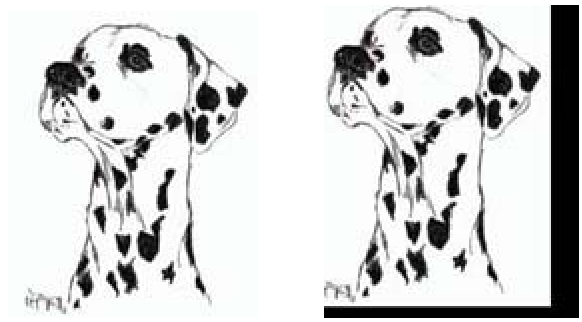
Dithering of the image to add noise to the image:
The images in the dataset are dithered in specific way that has performed in order to train the model to detect and classify the image even when there is noise.



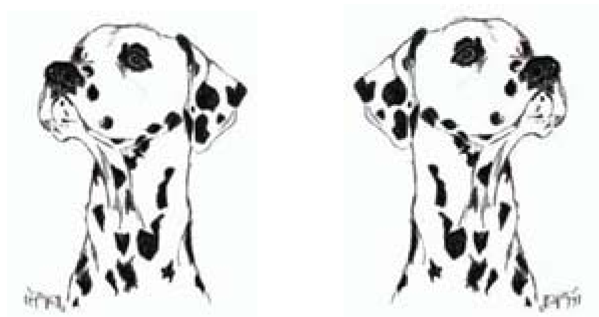
Creating Primary Colors Version of the image:
The images in the dataset are converted to primary colored images. The primary colors are the red blue and green so, for any specific pixel, we take a threshold value of 127 and if the value of red, green or blue is greater than that then, it you assign a value of 255 to the same and 0 in all other cases to that specific value of R, G and B respectively.



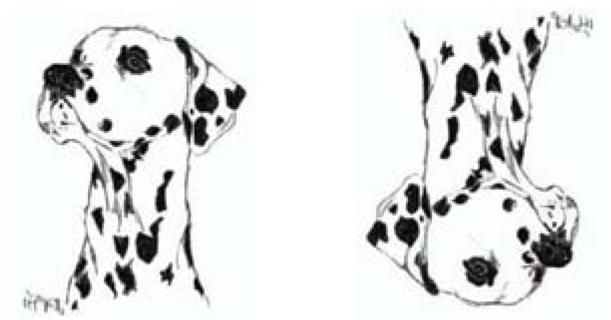
Random translation of the image:
Every image in the dataset is translated or moved by random number of pixels ranging from 0 to 30. These pixels are colored black making it a totally different image for the model as it captures different feature this time.



Flipping the image Vertically:
Flipping the image vertically using the user defined function as shown below.



Flipping the image Horizontally:
Flipping the image horizontally using the user defined function as shown below.



RESULTS OF CNN-

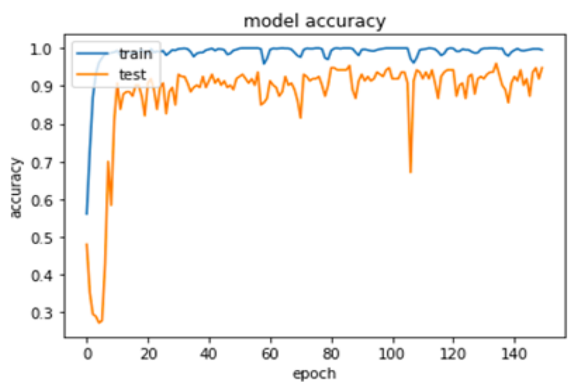


Figure 5: Graph of Variations in Accuracy with Epoch

Authorized licensed use limited to: University of Southern Queensland. Downloaded on July 04,2020 at 16:48:17 UTC from IEEE Xplore. Restrictions apply.

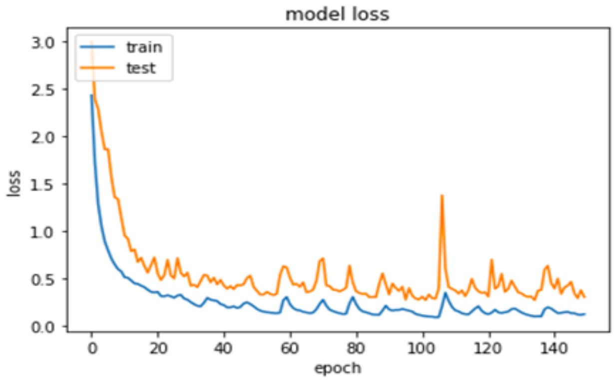


Figure 6: Graph of Variations in Loss with Epoch

A row in the confusion matrix shows the number of images that were of some class and the predicted class. All the numbers in the main diagonal shows the number of correct predictions by the model.

	Dalmatian	Dollar Ball	Pizza	Soccer Ball	Sunflower
Dalmatian	130	0	12	3	6
Dollar Ball	0	125	3	1	3
Pizza	1	0	199	1	6
Soccer Ball	3	0	16	233	2
Sunflower	4	0	8	1	182

	Dalmatian	Dollar Ball	Pizza	Soccer Ball	Sunflower
Precision	0.94	1	0.83	0.97	0.91
Recall	0.86	0.94	0.96	0.91	0.93
F1-Score	0.89	0.97	0.89	0.94	0.92

Figure 7: Accuracy Assessment for CNN

The above table clearly shows that the accuracy of CNN model has been drastically increased from that of SVM model.

VI. CONCLUSION AND FUTURE WORK

In conclusion, we implemented SVM using a very small dataset to achieve an accuracy of 93% and although SVM is a very strong technique, achieving such a high accuracy is still an anomaly. Eventually, we realized that our results boasted such a high accuracy due to the lack of a large enough dataset. So, using data augmentation, we more than tripled the size of our dataset and on performing SVM again, we achieved an accuracy of 82%, a significant decrease. Unsatisfied with the results, we decided to move to other deep learning techniques. This quest led us to Neural Networks and, CNN. On successfully implementing CNN, we achieved an accuracy of a staggering 93.57% on the very same dataset. This stands as a testimony to the increased potential of deep learning techniques over the more traditional machine learning techniques. Convolutional Neural Networks (CNNs) are similar to the more ordinary neural networks in that they are made up of hidden layers consisting of neurons with learnable parameters and conventionally, the SoftMax function is the classifier used at the last layer of this network. However, our research noted that there have been many studies that say otherwise. These studies introduce the usage of Support Vector Machine (SVM) in an artificial neural network architecture. Now that we have the knowhow to implement both CNN and SVM, we think that we will be able to support or challenge these studies and if we do decide to take this project further, this is the direction we will be heading towards.

VII. ACKNOWLEDGEMENT

I would like to acknowledge with gratitude, the support and love of my professors, and they all kept me going and this research would not have been possible without them.

REFERENCES

1. Pooja Kamavisdar et. Al A Survey on Image Classification Approaches and Techniques, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2013.
2. PriyaPradipNaswale & P. E. Ajmire,Image Classification Techniques-A Survey, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 5, Issue 2, March-April 2016
3. Chaitali Dhaware et. Al, Survey On Image Classification Methods In Image Processing, International Journal of Computer Science Trends and Technology (IJCS T) – Volume 4 Issue 3, May - Jun 2016.
4. Nadia Jmour et. Al Convolutional neural networks for image classification, International Conference on Advanced Systems and Electric Technologies (IC_ASET),14 June 2018.
5. M. Manoj Krishna et. Al Image classification using Deep learning, International Journal of Engineering & Technology, 7 (2.7) (2018) 614-617.
6. Krizhevsky, A., Sutskever, et. Al ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25 (pp. 1106-1114).
7. Le Hoang Thai et. Al Image Classification using Support Vector Machine and Artificial Neural Network, I.J. Information Technology and Computer Science, 2012, 5, 32-38, May 2012 in MEC, DOI: 10.5815/ijitcs.2012.05.05.
8. Maneela Jain et. Al, Review of Image Classification Methods and Techniques, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 Vol. 2 Issue 8, August – 2013.

Authorized licensed use limited to: University of Southern Queensland. Downloaded on July 04,2020 at 16:48:17 UTC from IEEE Xplore. Restrictions apply.