A Dissertation Report on

Image Classification using Logical Analysis of Data

Submitted for the partial fulfilment of the requirements for the award of

course credits for

Master of Technology

Submitted by

Tejas Deshpande

M. Tech. CSE

Enrollment No: 17535005

Under the guidance of

Dr. Balasubramanian Raman

Professor

Department of Computer Science and Engineering

Indian Institute of Technology, Roorkee

Roorkee – 247667

May, 2019

AUTHOR'S DECLARATION

I declare that the work presented in this dissertation with title "**Image Classification using Logical Analysis of Data**" towards fulfilment of the requirement for the award of the degree of **Master of Technology** in **Computer Science and Engineering, Indian Institute of Technology Roorkee , India** is an authentic record of my own work carried out during the period of **June 2018 to May 2019** under the supervision of **Dr**. **Balasubramanian Raman** , Professor, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, India. The content of this dissertation has not been submitted by me for the award of any other degree of this or any other institute.

Date: Tejas Deshpande

Place: Roorkee (17575005)

CERTIFICATE

This is to certify that Thesis Report entitled "**Image Classification using Logical Analysis of Data**" which is submitted by Tejas Deshpande (17535005), towards the fulfilment of the requirements for the award of the degree of **Master of Technology** in **Computer Science and Engineering,** submitted in the **Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, India** is carried out by him under my esteemed supervision and the statement made by the candidate in declaration is correct to the best of my knowledge and belief.

Date:

Place: Roorkee

Sign:

Dr. Balasubramanian Raman

Professor,

Indian Institute of Technology Roorkee.

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. Balasubramanian Raman for guiding me throughout my thesis work, helping me whenever needed and being a constant source of motivation. I am really grateful for having such a wonderful and understanding mentor. I am also thankful to the Department of Computer Science and Engineering, IIT Roorkee for providing the valuable resources to aid my research. I would like to thank my fellow classmates and friends for being there and helping me overcome the obstacles in my thesis work.

Last but not the least, I would like to thank my parents and my sister for their blessings and support without which I would not have reached this stage of my life.

Tejas Deshpande

ABSTRACT

Convolutional Neural Networks (CNNs) dominate various computer vision tasks since Alex Krizhevsky showed that they can be trained effectively and reduced the top-5 error from 26.2 % to 15.3 % on the ImageNet large scale visual recognition challenge. Many aspects of CNNs are examined in various publications, however, CNNs come with a set of disadvantages and limitations. CNNs or any deep learning models have no interpretability by humans. The reasoning behind a prediction in a CNN can never be understood and that is a problem when building reliable and robust AI solutions. Another issue with the CNNs is that they require huge amounts of data to work well. Feature extraction from images has been a popular technique in many computer vision problems. This report proposes a solution based on feature extraction from techniques like SIFT, SURF, HOG, etc. and using representation learning for learning different features from training images. Logical Analysis of Data is then used to classify images from one class label to another to solve the interpretability problem.

Contents Page Number

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image classification has been an interesting area of research for a while. A lot of work has been done to improve image classification accuracy till now. The formal definition of the problem of image classification is – Given an image, find out the label of the most prominent object in it. Computers do not have a visual system that enables them to decide what a particular image is of. An image is stored as a two-dimensional matrix having the intensity of colour for each position in the image. It is not possible to find out the contents of the image by just looking at the numbers. Factors such as scaling, rotation, viewpoint or illumination can cause huge changes in the matrix even though two images have the same subject. Good image classifiers should classify images based on the prominent object's general shape, size or colour. Over the years, various sophisticated algorithms have been proposed for the image classification problem. Feature Descriptor based algorithms like SIFT [1] and SURF [2] first appeared around 1990's and were the most popular algorithms till the 2010. These algorithms relied on finding a few interesting points in the image that can be used as features to describe these objects. These key points are represented in such a way that they are immune to scaling, rotation, illumination and viewpoints. These key points need to be easily recognisable in various similar images and therefore they are generally points in the images whose immediate neighbours are in contrast with them. Traditional classifier techniques like k- Nearest Neighbours and Support Vector Machines [3] were used to actually classify the images based on these features.

The early years of 2010s saw a boom in the Neural Network Architectures. A large number of images were easily available due to the internet and the computing powers of the machines had increased tremendously. Millions of labelled images were available now for use. Hardware specifications of GPUs also enabled to perform parallel computations efficiently. Neural network architectures work remarkably well when huge amounts of data is available. A novel technique for classifying images had just been introduced in the ImageNet 2012 Challenge. Thus, a series of new and improved neural network architectures were published in the following years. Each subsequent new solution was built on top of the AlexNet and the models are becoming deeper and deeper with each new solution. Various methods have been proposed that take advantage of the statistical constraints and relations between the data that have resulted in increased accuracy or reduced computations. The accuracy for most of the image classification models based on deep learning is close to or on par human performances and very little error rate is remaining. Deep learning models can now identify an image's subject very accurately. Now the thing that they cannot do is to do it computationally cheaply.

Logical Analysis of Data (LAD) is a classification technique that has been used in medical fields since the late 1980s. LAD has been instrumental in creating automated systems for predicting if a person would have a particular disease based on their symptoms and other parameters like height, weight, results of certain measurements, the expression levels of genes or proteins in the blood of the patients[4]. LAD is a data mining technique that classifies based on pattern recognition. LAD is applied in two stages: training phase, and a testing or theory formation phase, in which part of the database is used to extract special features or patterns of some phenomena, and the rest of the database is used to test the accuracy of the previously extracted knowledge. It is a supervised learning algorithm. this means that the database contains its class labels. LAD was first proposed in 1988. After many years, LAD become one of the most promising data mining methods developed to date for extracting knowledge from data [5]. LAD is based upon partial Boolean functions which take into account a fraction of all the possible configuration states when creating a rule-based system for classification. LAD is a non-statistical approach, thus there is no need to make certain assumptions regarding the posteriori class probabilities. This is an attempt to use Logical Analysis of Data as a classifier for images.

1.2 Problem Statement

Image Classification using Logical Analysis of Data

In the recent years, various deep learning models have been used for solving the image classification problem. The accuracy deep learning models provide is extremely high, however deep learning models come with their own set of problems. The computational cost of training a deep learning model is extremely high. Also, these models are not able to justify their decisions so that humans can review them. There is no information that is interpretable by humans that can be read and understand the reasoning behind the predictions made by these models.

The issue of interpretability / justification can be solved by using Logical Analysis of Data (LAD) for classification. LAD classifies data having attributes available in tabular format. Various feature extraction techniques can be used to select the important features from a image and then convert them into the tabular format. Also, it has been shown that LAD works better than deep learning models when the available training dataset is not enormous. This report further expands on these points in detail.

CHAPTER 2

RELATED WORK

2.1 Literature Review

One of the first Image classification algorithms were proposed in 1990's which relied on finding some interesting points in the images to compare them with the interesting points extracted from the training images. A lot of images with their labels are provided to the model. The model then applies the feature extraction algorithm to each of the images and collects a set of interesting points for each class. In the testing phase, a test image is provided and the same feature extraction process is run on this image. The decision is taken based on the similarity of the interest points with respect to each class interest points.

Scale Invariant Feature Transform (SIFT) was introduced in [1]**.** The Scale Invariant Feature Transform is invariant to scale, viewpoint, illumination and rotation. These invariance to all of the mentioned attributes made SIFT very accurate. SIFT deals with the scaling aspect of object by taking the input image and scaling it multiple times. Candidate key points are generated by taking Difference of Gaussian of each pixel with respect to its various scaling. These candidate key points are further reduced by considering only the points which have higher difference than all of its neighbours. Various low contrast points and points along the edges are removed. In the end features for the images independent of scale, illumination, rotation and perspective are found.

Another interest point based approach was introduced in [6]. The Histogram of Oriented Gradient attempts to find corners by using the gradient of the image. These gradients of each points with respect to their neighbours is aggregated into different bins of gradients. A set of these points' features is used as the basis of differentiating images from one another.

Algorithms introduced in [1] and [6] were used to represent a given image into a vector of fixed sizes. For all the images in the training data, a number of interesting points were found for each image. Now for all the interesting points, clustering algorithms like K-Means or DB-Scan are used to get clusters of interesting points. These clusters are considered for checking similarity of the new testing images. This creates a robust system having features that are the aggregate of all the training data. On test data, the key points generated by the feature extraction algorithm are compared to the cluster centres found out using the clustering algorithm to determine the closeness of each key point to a previously seen point. A classifier can now classify the image into one of the classes based on the training data. These kinds of systems have been the most popular methods for image classification till the 2010s. With the rise of available data through the internet and having enough computing power to work with millions of images, new set of algorithms surfaced.

The ImageNet challenge is a competition held every year for finding out the best image classification algorithms. The dataset of ImageNet consists of around fourteen million images of around a thousand different classes. Each of the image has been hand annotated. The criteria for the ImageNet challenge is that the participant's model should output five different classes that an image could be classified as, and the result is considered correct if the ImageNet's provided label is present in the participant's five class labels.

In 2012, the first neural network-based architecture was introduced [7]. ”Alex Net”, the convolutional neural network architecture had a top-5 error rate of 15.3% while the SIFT-like classification algorithm had an top-5 error rate of 26.2%. The huge increase in accuracy paved a way for various new and improved CNN Architectures in the future. The architecture had five convolutional filters, max- pooling layers and three fully connected layers.

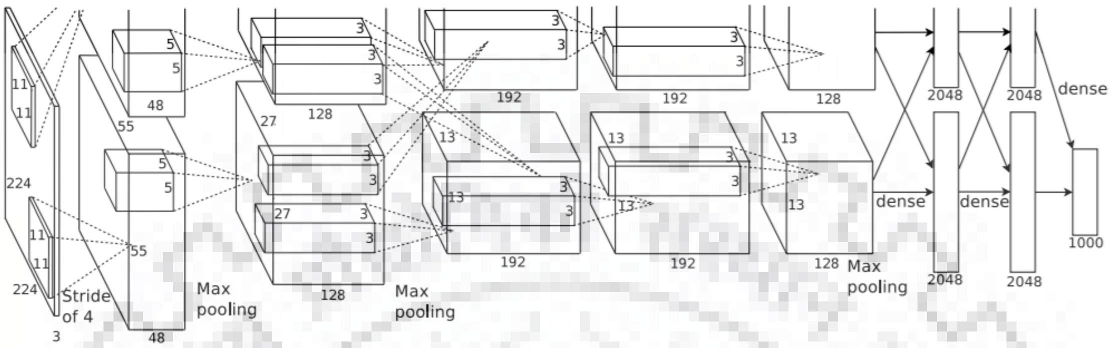

Figure 1: Alex Net CNN Architecture [7]

2014 saw another drop-in top-5 error rate with the introduction of [8]. The VGG16 model consists of sixteen convolutional layers, max-pooling and three fully connected layers. This model was much deeper than its predecessor having sixteen layers as compared to three. Also, 3x3 filters were used instead of Alex Net’s 11x11 which had the capability of learning the same patterns but had considerably less trainable parameters. The top-5 error rate of VGG16 was 7.3%.

In the same year, researchers have developed the concept of inception modules. Original convolutional layer used linear transformations with a nonlinear activation function. However, training multiple convolutional layers simultaneously and stacking their feature maps linked with a multi-layer perceptron also produces a nonlinear transformation. This idea has been exploited by [9] , who proposed a deeper network called GoogLeNet, also known as Inception V1 with 22 layers using such “inception modules” for a total of over 50 convolution layers. Each module is composed of 1x1, 3x3, 5x5 convolution layers and a 3x3 max-pool layer to increase sparsity in the model and obtain different type of patterns. The feature maps produced are then concatenated and analysed by the next inception module. The GoogLeNet model has a 6.7% error rate over the 2014 ImageNet challenge which is somewhat lower than the VGG16 but astonishingly smaller size (55 MB vs 490 MB). This gap is mainly due to the presence of the three large fully-connected layers in the VGG architecture.
In 2015, researchers in [10] developed the Inception V2 model, which was inspired by InceptionV1. The 5x5 filter was however replaced by two 3x3 filters, a 3x3 convolution and a 3x1 fully-connected layer slide over the first one. This method decreases the number of parameters in each inception module, thus reduces the total computational cost. InceptionV2 had a top-5 error rate of 5.6% on the 2012 ImageNet challenge.

Also, a fine-tuned batch-normalization process has been introduced in [11], and used a higher resolution input. Reduction of the strides of the first two layers and removal of a max-pool layer to analyse images with higher precision is the key change.

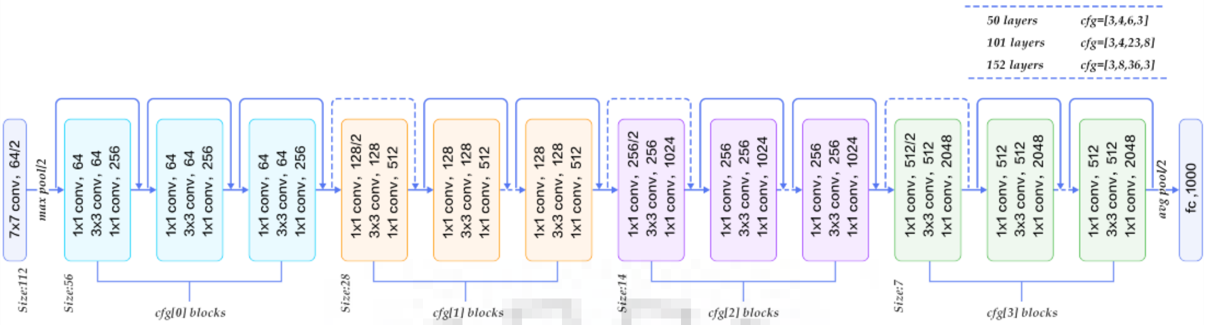

Figure 2: ResNet Architecture [12]

The top-5 error rate achieved using this is 3.58% over the 2012 ImageNet challenge. Residual Learning models have reduced the error rate to 3.57% [12].

2.2 Research Gaps

Although deep learning models have been successful in increasing image classification accuracies in the past, they have all stagnated at the top-5 accuracy of around 97% on the ImageNet Challenge. Various efforts in creating a deeper model have been made, however, there has not been a significant increase in accuracy compared to the previous models. This is mainly because, deeper models have huge amounts of trainable parameters which do not optimize well due to their numbers and contribute to the error rate. Various techniques have been introduced that work very well for domain specific images, however these techniques fail to generalize for a general dataset. In the recent years, there have been approaches that try to minimize the number of trainable parameters which can give equal or better quality of features. Alternative approaches can lead to better results.

One major drawback of neural network architectures is that the reasoning behind the decisions made by them is not interpretable by humans. Although deep learning models have high accuracy on most of the task, they are only suitable for research purposes. Applications of neural network models in real world cannot be used until we make sure that no decision is made that can severely affect a human's life adversely. Neural networks do not have the concept of support sets which can be tracked to identify the frequency of occurrence of a particular pattern. In the real world, where actions taken based on the predictions of AI systems affect everyone, cannot and should not be taken without knowing the reason behind the prediction. Sectors such as aviation, military applications cannot use the deep learning models as there is no guarantee of the decisions taken by these models. The lack of justifiability of decisions is an important drawback of the current deep learning models. Also, due to lack of interpretability by humans, neural networks can be difficult to debug. A source of bug may never be understood and the whole model may have to be reimplemented / redesigned because of this. A lack of interpretability hinders development process for building solutions.

One aspect in which deep learning models lack is the cost to develop a solution. Deep learning models are computationally very expensive. The time required for training a neural network model is tremendous. It is not unheard of to train a neural network for several weeks to get a good accuracy rate. The AlexNet model needs 14 days of training without GPUs. This type of process is only viable for specific circumstances. Faster and cheaper models need to be used to get a quick prediction while waiting for the deep learning model to finish training. Traditional classifiers and data mining techniques can be trained to classify images in matter of hours. The difference in cost of time and resources is huge to not consider the traditional models especially when having the optimal accuracy is not the top priority.

The demand for data in deep learning models is huge. Millions of images have to be provided to the model in order to achieve good results. Although sometimes deep learning models work with limited data, most of the times they do not. The demand for such huge amounts of data is tremendous and cannot always be met. For smaller datasets, traditional machine learning models work much better than deep learning models. Furthermore, when having a dataset which requires domain-specific knowledge, deep learning models do not work very well. Feature engineering using domain expert’s knowledge is required to get better results. However, if the classification algorithm is pattern based, then the classifier can extract these features automatically. This is be of huge help as manual intervention is not needed in such cases. LAD has been used in medical research extensively and produces good results. However, it has not been implemented as an image classifier before. Various feature extraction algorithms can be used in conjunction with LAD to solve the problem.

A few of the above discussed problems are addressed in the proposed solution.

CHAPTER 3

PROPOSED SOLUTION

3.1 Feature Extraction

Feature Extraction is the process of extracting a few interesting points from an image and to represent them in a way that can be interpreted by a classifier. These interesting points should be used to distinguish an object of one class from another. Various feature extraction features have been suggested in the field of Computer Vision. The SIFT algorithm caused a revolution in the early 2000's. SIFT finds various localized features that can be used to compare different images for similar content.

SIFT descriptors are multi-image representations of an image neighbourhood. They are Gaussian derivatives computed at 8 orientation planes over a 4x4 grid of spatial locations, giving a 128-dimension vector. SIFT generates features that are immune to scaling, rotation, illumination and viewpoint. Hence, they provide fairly robust representation of the key points.

Figure 3: SIFT based Matching

Top Left: Four template images   Top Right: Test Image

Bottom: Results of SIFT Matching [1]

Moreover, the SIFT algorithm runs extremely fast and the computational cost associated with it is low. SIFT features can be used with Bag of Visual Words like Representation to find the most prominent features across all training images of a particular class label. Instead of SIFT, similar techniques like SURF and Histogram of Oriented Gradients(HOG) can also be used. These fixed sized descriptors can then be fed to the classifier for predicting its class label.

A second approach to solve this problem is to use deep learning to generate features from the images[14]. This field of study is known as Representation Learning and is quite popular these days. The main idea behind representation learning is to use a group of filters that can distinguish one class of images from the others. This can be achieved with the help of supervised algorithms or unsupervised algorithms. Multiple Layer Perceptron can be used to learn a different set of filters on each layer and is a supervised algorithm. Auto Encoders are another way of representing compressed information about images. The auto encoders try to learn representation of given set of instances with lowest possible construction error. These features can then be used to generate tabular features which are required for LAD classifiers.

If the feature extraction methods like SIFT, SURF and HOG give better accuracy, then the problem of computational complexity can be solved. However, the computational complexity will not be less when using Representation Learning. Representation Learning is not pursued further in this document.

3.2 Clustering

Vocabulary is a way of creating a feature vector for classification that maps the descriptors in query images to descriptors seen previously during training phase. One extreme of this approach would be to compare each query descriptor to all training descriptors: this seems impractical given the huge number of training descriptors involved (hundreds of thousands). Another approach would be to try to identify a small number of large clusters that are good at distinguishing a given class: for instance, some approaches operates with 6 parts per category. The best trade-offs of accuracy and computational efficiency are obtained for intermediate sizes of clustering.

Given all the features from the training samples, the goal is to find some clusters that can be treated as a similar group of points. The k cluster centres will be the representations of all the thousands of descriptors from training. Whenever new features of testing images arrive, these features are compared to the k cluster centres to find the similarity between the two.

We chose to use the square-error partitioning method: k-means. This algorithm proceeds by iterated assignments of datapoints to their closest cluster centres and re-computation of the cluster centres. Two problems are that the k-means algorithm converges only to local optima of the squared distortion, and that it does not determine the parameter k. The value of k was determined by checking for various values and keeping the ones that provide good results.

3.3 Classifier

Logical Analysis of Data (LAD) has been effectively used for many classification problems before. LAD has achieved more accuracy than neural networks and other machine learning algorithms in many cases [13]. LAD seems to work very well for medical data and has been extensively used in AI systems in healthcare industry. LAD is based on the concept of pattern generation and support sets of patterns which classify data item based on the patterns mined from it. The support set provides interpretability to humans and hence this type of models' decisions can be justified. Given a training dataset, the model will find various positive and negative patterns. Positive patterns are a set of conditions that have been fulfilled only in the data items belonging to a positive class and have never been fulfilled in the data items belonging to the negative class. LAD becomes a powerful tool for classification when the features available are in tabular format. However, for the image classification problem, the features are not readily available in tabular data. Various interesting points from the image have to be extracted and converted into tabular format. LAD solves the issue of interpretability and justifiability of decisions.

There are three steps involved in training images for LAD. The first step is binarization which converts a real-valued attribute into multiple binary attributes. The second step involved tries to reduce the number of attributes by discarding the irrelevant ones. This process is called support set minimization. The last step is to generate patterns for each of the class. Whenever a testing data point arrives, check for each class if their patterns cover this data point. If it does, then the data point is classified into that particular class. These methods are described in more detail below.

3.3.1 Binarization

As LAD is based on partial Boolean functions, all the attributes need to be Boolean in the dataset. Binarization is the technique of converting a real-valued attribute into several binary attributes. Each of these binary attributes takes the value 1 if the real-valued attribute to which it is associated takes values above a certain threshold. Similarly, they take a value of 0 if the real-valued attribute take value below the threshold.

The cut points are chosen in a way which will allow to distinguish between positive and negative classes. If we introduce two cut points between the consecutive values of a certain attribute, then the corresponding binary variables will be equivalent, and therefore will not help us in any way to make a distinction between the points in positive class and negative class. Similarly, cut points above the smallest and the largest entry will not provide any help in distinguishing points of positive and negative class. Therefore, we shall consider at most one cut point between any two consecutive values of attribute A, and no cut points above the largest entry and the smallest entry.

Firstly, the attribute to be binarized is sorted in ascending order. If for two adjacent entries the class labels corresponding to the two entries are different, then a new threshold is introduced which would be the average of the two entries. If two adjacent entries belong to the same class, then no threshold is introduced to separate them into different binary variables.

Attributes	A	B	C
Class 0	3.5	3.8	2.8
	2.6	1.6	5.2
	1.0	2.1	3.8
Class 1	3.5	1.6	3.8
	2.3	2.1	1.0

Table 1: A numerical dataset before binarization [16]

Attributes	A			B		C		
Variables	A1	A2	A3	B1	B2	C1	C2	C3
Class 0	1	1	1	1	1	0	0	1
	0	1	1	0	0	1	1	1
	0	0	0	0	1	0	1	1
Class 1	1	1	1	0	0	0	1	1
	0	0	1	0	1	0	0	0

Table 2: A numerical dataset after binarization [16]

3.3.2 Support Sets

After converting the real-valued attributes to binary attributes, the total number of attributes increase largely. Getting rid of the attributes that are irrelevant is important in order for the classifier to be trained in a realistic time. Support sets play an important role in minimizing the number of attributes that are considered when deciding if a particular test image belongs to that particular class.

A support set of a particular class is a set of attributes/variables that are needed to classify all the datapoints to that class. A support set is minimal if removal of any attribute from it fails to cover only the one class it belongs to. This problem is equivalent to the famous set-covering problem. However, in order to get more robust support sets that cannot be influenced by minor things, it is always a good idea to use a variant of the set-covering problem which is d-set covering problem. Here, d is the frequency of the specific configuration of the attributes corresponding to that particular class. The value of d can range from 3 to even 20 depending upon the dataset in hand. If there are a lot if variables and training examples, then higher values of d are used and if the attributes and the training examples are less, then lower values of d are used.

3.3.3 Patterns

The pattern generation procedure is an important step in the LAD classifier. A pattern for a particular class is defined as a conjunction of some attributes, which is observed to be true for at least one observation of that class, and false for all observations of other classes in the training data set. The degree d of a pattern is the number of binary attributes used to define the pattern. If a pattern covers a datapoint, this means that this observation is from the same class to which the pattern belongs. There are many techniques for pattern generation, for example enumeration, heuristics, and linear programming. It has been observed that simple decision tree algorithms work quite well after binarization and support set minimization step.

CHAPTER 4

DATASET USED

4.1 CIFAR-10

The CIFAR-10 dataset is a collection of 60,000 images having resolution of 32 x 32 pixels. The CIFAR-10 dataset contains colour images in 10 different classes. The 10 different classes represent airplanes, birds, cars, cats, deer, dogs, frogs, horses, ships, and trucks. There are 6,000 images of each class. The dataset was collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton from the University of Toronto [17]. The dataset consists of five training files each of which contains 10,000 different images. All the training images are randomly distributed across the five files, so each file may have different number of images per class. Finally, there is one testing file which also contains 10,000 images.

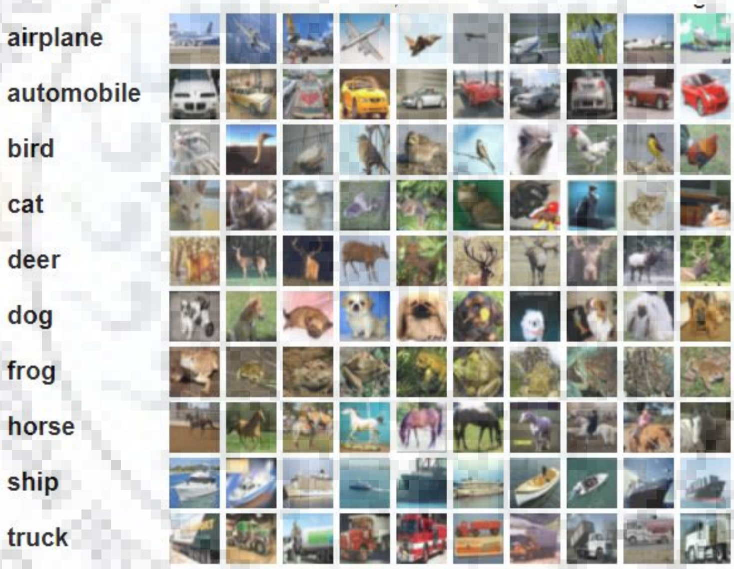

Figure 4: CIFAR-10 Dataset [17]

4.2 CIFAR-100

The CIFAR-100 dataset is similar to the CIFAR-10 dataset. The dataset consists of 100 class labels and 20 superclass labels. Each superclass has five different classes which are similar visually or semantically to each other. Each class has 500 training images and 100 testing images. Each image is of size 32 x 32 pixels [17].

4.3 Caltech101

Caltech101 has images of objects belonging to 101 categories. There are about 40 to 800 images per category. Most classes have about 50 images. The size of each image is 300 x 200 pixels. This dataset was collected in September 2003 by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato. [18]

CHAPTER 5

RESULTS

Using LAD with SIFT descriptors gives descent results when compared to traditional learning models. When creating the vocabulary with k-Means clustering the value of k needs to be set manually. It was noticed that changing the value from 50 to 200 didn't change the accuracy significantly on the datasets. However, having higher values of k increases the running time of the program slightly. Hence, value of k=100 was used for reporting the results.

Regarding the time taken to train the model, LAD takes a significant time to train with respect to SVM. Training images from the CIFAR-10 dataset took roughly 5 hours for SVM whereas it took 9 hours to train LAD. During the testing phase however, LAD performed on par with SVM classifier. Using Fischer Vectors (FV) with SVM have shown to increase their accuracy by quite a much in [19]. Fischer Vectors provide a greater degree of information about the cluster representatives and the individual datapoint's closeness to other cluster representatives. Thus, FVs give a richer set of features for training thus increasing the accuracy.

When training on the Caltech101 dataset, the standard procedure of training on 30 images and testing on the remaining is used. A random forest approach has been suggested in [20] [21] which has been included in the comparison. Below are the comparisons of various approaches for the three datasets.

1.CIFAR-10

Approach	Accuracy
SIFT+ SVM	71.86%
SIFT + SVM + FV	79.59%
SIFT + LAD	76.31%
AlexNet	89.34%

Table 3: Comparison of approaches used for CIFAR-10 Dataset

2.CIFAR-100

Approach	Accuracy
SIFT+ SVM	66.55%
SIFT + SVM + FV	73.62%
SIFT + LAD	72.02%
AlexNet	81.64%

Table 4: Comparison of approaches used for CIFAR-100 Dataset

3.Caltech-101

Approach	Accuracy
SIFT+ SVM	81.30%
Random Forests	80.00%
SIFT + LAD	79.83%
AlexNet	92%

Table 5: Comparison of approaches used for Caltech-101 Dataset

CHAPTER 6

CONCLUSION

Various feature extraction techniques have been discussed. Techniques like SIFT, HOG, etc. show a lot of promise for representing images. Representation learning is another approach worth trying. Combining these techniques with Logical Analysis of Data as a classifier, many of the discussed issues can be solved. LAD has proven to be quite useful in the field of medicine and healthcare. Image classification can be a new area where it would be very useful. The accuracy of using LAD seems to be on par with other classifiers like SVM and Random Forest Classifier. However, much more can be achieved in terms of time taken for training. Various combinatorial optimizations have been suggested for LAD in the past that may speed up the training process [22]. Also, other methods for pattern generation can be developed that can improve the quality of patterns in the classifier. However, using LAD as a classifier is definitely a viable solution when working under limited time and computational resources.

REFERENCES

[1] David G. Lowe. “Object recognition from local scale-invariant features”. In *Computer Vision and Pattern Recognition, 1999, pp.* 1150-1157.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. “Surf: Speeded up robust features”. In *European Conference on Computer Vision*, 2006

[3] C. Cortes, and Vapnik, “Support Vector Networks”. In *Machine Learning Journal*, 1995, Volume 20, Issue 3, pp. 273-297.

[4] Boros, E., P.L. Hammer, T. Ibaraki and A. Kogan. “An Implementation of Logical Analysis of Data”. In *IEEE transactions on knowledge and data engineering*, 2000, pp. 292-306.

[5] Han, J., Kim, N., Yum, B.-J., & Jeong, M. K. “Pattern selection approaches for the logical analysis of data considering the outliers and the coverage of a pattern”. In *Expert Systems with Applications*, 2011.

[6] Navneet Dalal and Bill Triggs. “Histograms of oriented gradients for human detection”. In *Computer Vision and Pattern Recognition, 2005.*

[7] Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. “ImageNet classification with deep convolutional neural networks”. In *Advances in Neural Information Processing Systems 25*, 2012, pp. 1106–1114.

[8] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In *Computer Vision and Pattern Recognition*, 2014, pp.1556.

[9] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. “Going deeper with convolutions”. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2015, pp. 1–9.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” In *Conference on Computer Vision and Pattern Recognition*, 2016.

[11] Ioffe, S., Szegedy, C. “Batch normalization: Accelerating deep network training by reducing internal covariate shift” *, 2015.*

[12] He, K., Zhang, X., Ren, S., Sun, J.: “Deep residual learning for image recognition”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. pp. 770– 778.*

[13] Hammer, P.L. & Bonates, “Logical analysis of data—An overview: From combinatorial optimization to medical applications” In *Annals of Operations Research*

November 2006, Volume 148, *pp.* 203–225.

[14] Baldi, 2012, “Autoencoders, unsupervised learning, and deep architectures” In *Journal of Machine Learning Research, 27 ,2012, pp. 37-50.*

[15] Mohamed Hamdi-Cherif, “Logical Analysis of Data (LAD) - Usage and Experimentation” In *proceedings of the International Conference on Artificial Intelligence, 2017, pp. 266-272.*

[16] G. Alexe, S. Alexe, T.O. Bonates, A. Kogan. “Logical Analysis of Data – the vision of Peter L. Hammer” In *Annals of Mathematics and Artificial Intelligence*, 2007, pp. 265-312.

[17] A. Krizhevsky. “Learning multiple layers of features from tiny images”. *Tech Report*, 2009.

[18] L Fei-Fei, R Fergus, P Perona. “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories” In *Computer vision and Image understanding*, 2017, pp. 59-70.

[19] J. Sánchez and F. Perronnin. “High-dimensional signature compression for large-scale image classification”. In *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on, pp. 1665–1672.

[20] A. Bosch, A. Zisserman, and X. Munoz. “Image classification using random forests and ferns”. In *International Conference on Computer Vision*, 2007, pp.1-8.

[21] L. Breiman. “Random forests”. In *Machine Learning*, 2001, Volume 45, Issue 1, pp. 5–32.

[22] M. Subasi, A. Herrera “Logical Analysis of Multiclass Data” In *Latin American Computing Conference (CLEI)*, 2015