CLASSIFICATION AND AESTHETIC EVALUATION OF PAINTINGS & ARTWORKS

A Project Report

submitted in partial fulfilment of the

requirements for the award of the degree of

Bachelor of Technology (B.Tech.)

Submitted by

TARPIT SAHU

(140060101108)

SONU KUMAR

(140060103056)

ARJUN TYAGI

(140060101022)

Under the supervision of

Mr.Neeraj Kumar Pandey
Assistant Professor
IT Department

Mrs.Nilima Patel
Assistant Professor
IT Department

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING ROORKEE
ROORKEE –247667 (UTTARAKHAND) INDIA

May, 2018

ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along the completion of my project. All that I have done is only due to such supervision and assistance and I would not forget to thank them.

I respect and thank Mr. Neeraj Kumar Pandey, for providing us an opportunity to do the project work in College Of Engineering Roorkee and giving us all support and guidance which made us complete the project duly. We are extremely thankful to him for providing such a nice support and guidance, although he had busy schedule managing the other affairs.

I owe my deep gratitude to Mrs. Nilima Patel, who took keen interest on my project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

I would not forget to remember our team mates for their encouragement and more over for their timely support and guidance till the completion of our project work.

I am thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs of Department of Computer Science & Engineering which helped me in successfully completing our project work. Also, I would like to extend our sincere esteems to all staff in laboratory for their timely support.

Tarpit Sahu

Arjun Tyagi

Sonu Kumar

CANDIDATE'S DECLARATION

We hereby declare that the work presented in this project report titled,**CLASSIFICATION AND AESTHETIC EVALUATION OF PAINTINGS AND ARTWORKS"** submitted by us in the partial fulfillment of the requirement of the award of the degree of **Bachelor of Technology (B.Tech.)** Submitted in the Department of **Computer Science & Engineering, College of Engineering Roorkee**, is an authentic record of our project work carried out under the guidance of Mr. Neeraj Kumar Pandey & Mrs. Nilima Patel, College Of Engineering Roorkee.

Date **Tarpit Sahu**

Place: Roorkee

Sonu Kumar

Arjun Tyagi

SUPERVISOR'S CERTIFICATE

It is to certify that the Project entitled **"CLASSIFICATION AND AESTHETIC EVALUATION OF PAINTINGS AND ARTWORKS"** which is being submitted by **Mr. Tarpit Sahu, Mr. Sonu Kumar, Mr. Arjun Tyagi** to the College of Engineering Roorkee in the fulfilment of the requirement for the award of the degree of **Bachelor of Technology (B.Tech.)**, is a record of bonafide project work carried out by him under our guidance and supervision. The matter presented in this project report has not been submitted either in part or full to any University or Institute for award of any degree.

Mr. Neeraj Kumar Pandey
Assistant Professor
IT Department

Mrs. Nilima Patel
Assistant Professor
IT Department

Mr. Himanshu Chauhan
Head Of Department
CSE Department

Mr. Brij Mohan Singh
Dean (C&D)

TABLE OF CONTENTS

Classification and Aesthetic Evaluation of Paintings and Artworks

Tarpit Sahu
CSE
COER
Roorkee, India
tarpitsahu@gmail.com

Arjun Tyagi
CSE
COER
Roorkee, India
arjuntyagi.t@gmail.com

Sonu Kumar
CSE
COER
Roorkee, India
sonu99803@gmail.com

Pushkar Shukla
CSE
University Of California
Santa Barbara, USA
pushkar_shukla@gmail.com

Abstract— **Painters and Artists have contributed to the field of art over the years with their exceptional talent and skills. The Internet is full of their creativity and imagination where one can find most of their work. Like any other information present on the Internet, paintings are also not well organized. In this paper, a method is proposed to classify paintings with the help of support vector machine classifier using features extracted by a pre trained convolutional neural network- AlexNet. A painting is not only an art on paper but is a medium to arouse emotions and sense of pleasure within the audience. Aesthetic Evaluation aims at evaluation/rating a painting or an artwork on the basis of various parameters like style, topic, emotional engagement etc. which cannot be done by a machine alone. So we cannot leave behind the human inputs while determining the aesthetic value of a painting or an artwork. In this paper we also propose a method to judge or evaluate the aesthetic value of a painting by training a regression model with several image features, like Local Binary Pattern for texture, color histogram for color, Histogram of Oriented Gradients for edges and GIST for scene recognition in the painting, against human ratings for each painting. A dataset constituting of 1225 digital images of paintings of 7 categories is used for classifying and evaluating the aesthetic value. The classification phase was found to have 92.73% accuracy and the evaluation phase performed with an accuracy of 64.15%.**

Keywords–Painting classification; Aesthetic Evaluation; Convolutional neural network; Histogram of oriented gradient; GIST

I. INTRODUCTION

In present Digital Era, the advancement in technology and increase in number of smart phones and computers has eased image retrieval to large extent. Now a user can easily search for an image by giving a simple query on his smart phone device. The concept of aesthetic evaluation will help in retrieving the images based on personal taste and choice on the basis of aesthetic scores.

The objective of the aesthetic assessment is to design methods which can automatically predict the perceived quality of a painting or an artwork. Such methods find great applications in the field of image retrieval. An automatic system that can evaluate painting's aesthetics has many potential applications. Using image retrieval systems, similar images could be ranked using aesthetic properties. They could help a user to select the best pictures from his collection to make photo albums. Also, these models could be deployed directly in photo cameras to make real-time suggestions and give on the spot scores to the photographs clicked. Knowing the aesthetic value of paintings will help in searching a painting and facilitate proper management of paintings. Using the proposed concept a recommendation application can be developed which will show the user a painting of required aesthetic quality. It will also be helpful in easing the process of searching a painting reducing both time and resources.

Figure 1. Painting categories used in this research.

Aesthetic Evaluation of a painting is a complex task. The evaluation not only depends on the theme and style of the artist but also depends upon the likings, disliking, personal taste and views of an individual evaluating the artwork. A piece of art which pleases some portion of a society need not necessarily please others. So we need to develop a system which can mimic the above-said concept efficiently. Fig. 1 shows the categories of paintings considered in this research.

II. LITERATURE REVIEW

Aesthetic Evaluation of images and photographs is a popular topic for research in Computer Vision field. There has been a lot of work done for assessing the aesthetic value of images and photographs, but there are few researches

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.39

done so far that deal with the assessment of aesthetic value of Paintings & Artworks.

AVA dataset [3] is one of the datasets widely used for assessing the aesthetic quality of an image. This large scale dataset containing more than 250,000 images was introduced by Murray et al. In their original work, they formulated a binary classification problem and established the experimental settings. They computed Fisher Vector signatures from SIFT descriptors and trained an SVM which achieved maximum accuracy of 67%.

Extensive research has been done on image classification and aesthetics using convolutional neural networks. Lu et al. [4] used convolutional neural networks on the AVA dataset and achieved classification accuracy between 60.25% and 71.2%. The architecture of convolutional network used by them contained 4 convolutional layers and 2 layers that were fully connected. Bianco et al. [5] used deep Convolutional Neural Network to predict image aesthetics. They fine tuned CNN architecture by casting the image aesthetic prediction as a regression problem and used AVA dataset. Their Experimental results show the robustness of the solution proposed, which outperforms the best solution in the state of the art by almost 17 % in terms (MRSSE). Marchesotti et al. [6] proposed to use generic image descriptors to assess aesthetic quality instead of hand-crafting features which would correlate with best photographic practices and achieved good results. Datta et al. [11] designed special visual features (colorfulness, the rule of thirds, low depth of field indicators, etc.) and used the Support Vector Machine (SVM) and Decision Tree (DT) to discriminate between aesthetic and unaesthetic images. Nishiyama et al. [7] proposed an approach based on color harmony and bags of color patterns to characterize color variations in local regions.

III. PROPOSED METHODOLOGY

The technique proposed for classifying digital images of paintings and evaluating the aesthetic value has been described in this section. Fig. 2 gives a block representation of the proposed approach to classify paintings. Fig. 4 illustrates flowcharts to evaluate aesthetic value of paintings. The dataset used in this research consists of 1225 paintings of 7 categories. The dataset was divided into training and testing data sets. 120 paintings of each category were used to train the algorithm whereas it was tested on 55 paintings of each category. The images were pre-processed as per the requirements of convolutional network.

To extract features of paintings for classification, a pre-trained Convolutional neural network- AlexNet [1] was used. AlexNet is one of the deep ConvNets which competed in the ImageNet Large Scale Visual Recognition Challenge in 2012. AlexNet has 5 convolutional layers, 3 sub sampling layers, 3 fully connected layers.

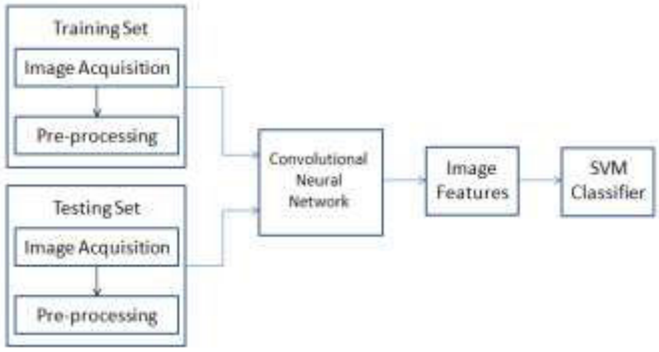


Figure 2. Block representation of proposed approach to classify paintings.

For evaluating the aesthetics of paintings, a survey was conducted and the participants were asked to rate the paintings on a scale of 1 – 10 where a rating close to 1 represents low aesthetic score and rating close to 10 represents high aesthetic score. Total 60 participants rated each of the 1225 paintings. A discussion with the participants revealed that people look for image features like colors, edges and texture in a painting. Keeping in mind the results of discussion following features of each painting were extracted: Colour Histogram, Histogram of Oriented Gradient, Local Binary Pattern and GIST.

A. *Histogram of Oriented Gradients*

HOG feature descriptors are widely used in Computer Vision for detecting edges. HOG features of the paintings are extracted and stored for further processing. HOG operates on local cells hence it is invariant to geometric transformations.

Figure 3. Extraction Of HOG Features
(a) Painting (b) Plot of HOG Features

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.39

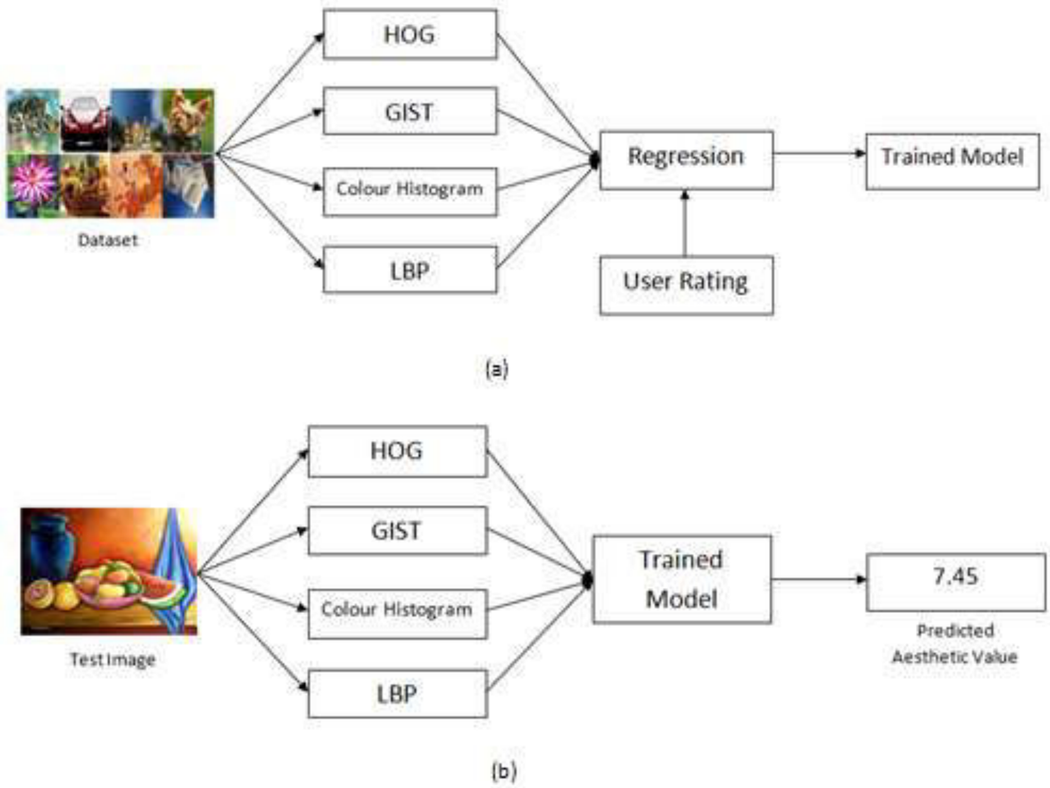


Figure 4. Block representation of proposed method to assess Aesthetic score:
(a) Training a Regression Model (b) Predicting Aesthetic Value using the trained model

B. *Local Binary Pattern*

LBP is a visual descriptor widely used in Computer Vision to classify images based on texture. Discriminative power and computational simplicity of LBP makes it a popular approach in various applications.

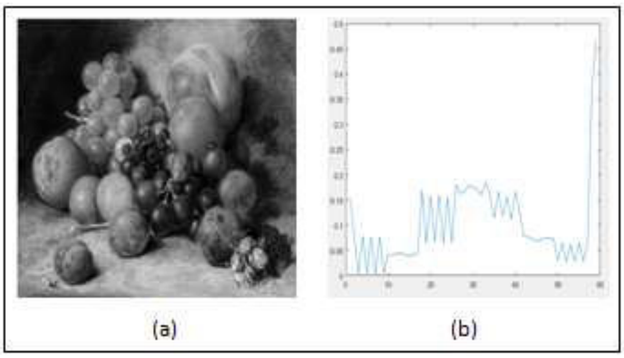


Figure 5. Extraction of LBP Features
(a)Original Painting converted to grayscale (b) LBP Histogram.

C. *GIST*

The GIST descriptor was first proposed for recognition of real world scenes the segmentation and the processing of individual objects or regions. To compute a GIST Descriptor, an image is convolved with 32 Gabor filters at 4 scales, 8 orientations, producing 32 feature maps. Each feature map is divided into 16 regions and the feature values of each region are averaged. Finally, the 16 averaged values of 32 maps are concatenated, giving a 512 GIST descriptor. Fig. 6 shows an example of GIST descriptor.

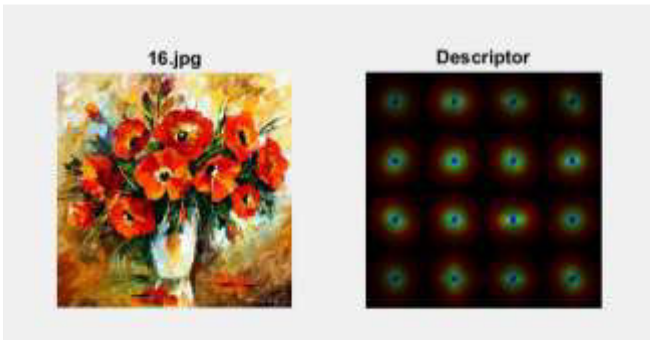


Figure 6. Extraction of GIST Features
(a)Original Painting (b) GIST Descriptor

After extracting the required features, a regression model was trained with image features as predictor variables and user rating as the response variable. Fig. 4 represents the block representation of proposed approach.

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.39

IV. EXPERIMENTAL RESULTS

This section describes the results of the proposed algorithms for classifying and assessing the aesthetic score of paintings. The results are described in two sections: first section describes about Classification problem and second section discusses about Evaluation problem.

A. *Classification of Paintings*

We used 2 layers ('fc6'&'fc7') of pre-trained AlexNet convolutional network for extracting the features of paintings. AlexNet extracted 8192 features (4096 by each layer). These features were then used to train an SVM classifier. Fig. 7 illustrates the confusion matrix obtained after using AlexNet and SVM for classification.

	A	B	C	D	E	F	G
A	**0.927273**	0.036364	0	0.018182	0	0.018182	0
B	0	**1**	0	0	0	0	0
C	0.072727	0	**0.836364**	0.054545	0	0.036364	0
D	0	0	0.054545	**0.854545**	0.054545	0.018182	0.018182
E	0	0	0	0	**0.963636**	0.036364	0
F	0.018182	0	0	0	0.054545	**0.927273**	0
G	0.018182	0	0	0	0	0	**0.981818**

A: Animal B: Car C:Flower D:Fruit
E:Monument F:Mountain G:Portrait

Figure 7. Confusion Matrix.

The model gave excellent results which is evident from the confusion matrix shown above. Paintings of Car, Portrait and Monument gave the best results whereas Paintings of Flower and Fruits gave results with very good accuracy. The proposed method obtained overall **92.73%** accuracy in classifying the paintings.

B. *Aesthetic Evaluation of Paintings*

After training a regression model, aesthetic scores of test data were predicted and compared with actual ratings given by the participants during the survey. Since prediction of exact aesthetic score of a painting is quiet difficult, following formula was used to determine the accuracy of our predictions:

If **| predicted_rating - average _rating | <= 1,** the prediction is said to be correct. Using described formula, our model was able to predict aesthetic scores of 247 paintings out of 385 test paintings according to the given range, thus giving an accuracy of **64.15%.** Fig. 8 illustrates the results obtained in this phase.

V. CONCLUSION AND FUTURE SCOPE

The experiments and data analysis in this project investigated a machine learning solution to aesthetic evaluation of paintings and artworks. The classification done, using CNN and SVM classifier, performed with an accuracy of 92.73% for over thousand paintings of seven categories. Evaluation of a painting is a complex task which sometimes depends on user's taste and understanding. Accuracy of 64.15% is achieved using image features like HOG, LBP, GIST etc combined with inputs from 60 persons on each painting. From the accuracy achieved it can be concluded that using image features and machine learning approach we can evaluate the aesthetic value of paintings. It is possible to improve the performance of the model by using more features and taking ratings from people of various domains. After improving the performance of the model it may be helpful in optimizing the search engines to show relevant images and paintings to the users. The search engine will show the results based on the aesthetic value input by the user. It would also be helpful in enhanced management of paintings. Art galleries and museums would be highly benefited by the aesthetic evaluation as it will help in attracting the right audience. More complex implementation can enable robots to automatically paint as per the requirement of the users or evaluate and provide guidelines to improve and enhance already existing artworks.

REFERENCES

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in neural information processing systems, pp. 1097-1105. 2012.

[2] Oliva, Aude, and Antonio Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." International journal of computer vision 42, no. 3 (2001): 145-175.

[3] Murray, Naila, Luca Marchesotti, and Florent Perronnin. "AVA: A large-scale database for aesthetic visual analysis." In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2408-2415. IEEE, 2012.

[4] Lu, Xin, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z. Wang. "Rating image aesthetics using deep learning." IEEE Transactions on Multimedia 17, no. 11 (2015): 2021-2034.

[5] Bianco, Simone, Luigi Celona, Paolo Napoletano, and Raimondo Schettini. "Predicting Image Aesthetics with Deep Learning." In International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 117-125. Springer, Cham, 2016.

[6] Marchesotti, Luca, Florent Perronnin, Diane Larlus, and Gabriela Csurka. "Assessing the aesthetic quality of photographs using generic image descriptors." In Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 1784-1791. IEEE, 2011.

[7] Nishiyama, Masashi, Takahiro Okabe, Imari Sato, and Yoichi Sato. "Aesthetic quality classification of photographs based on color harmony." In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 33-40. IEEE, 2011.

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.39

[8] Wu, Ou, Weiming Hu, and Jun Gao. "Learning to predict the perceived visual quality of photos." In Computer Vision (ICCV), 2011 IEEE International Conference on, pp. 225-232. IEEE, 2011.

[9] Y Kao, Yueying, Chong Wang, and Kaiqi Huang. "Visual aesthetic quality assessment with a regression model." In Image Processing (ICIP), 2015 IEEE International Conference on, pp. 1583-1587. IEEE, 2015.

[10] Datta, Ritendra, Jia Li, and James Z. Wang. "Algorithmic inferencing of aesthetics and emotion in natural images: An exposition." In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pp. 105-108. IEEE, 2008.

[11] R. Datta, D. Joshi, J. Li, J.Z. Wang “Studying aesthetics in photographic images using a computational approach.” In ECCV 2006. LNCS, vol. 3953, pp. 288–301. Springer, Berlin, Heidelberg, 2006.

Figure 8. Result of Aesthetic Evaluation. (a) Correct Prediction (Green) (b) Wrong Prediction (Red)

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.39

ABSTRACT

Painters and Artists have contributed in the field of art over the years with their exceptional talent and skills. The Internet is full of their creativity and imagination where one can find most of their works. Advancements in technology have eased image retrieval to a large extent; with one query to a search engine we can find paintings of any genre very easily but paintings, like any other information present on the Internet, are not organized. In this project, we illustrate amethod to classify paintings and artworks into various categories.

A painting is not only an art on paper but is a medium to arouse emotions and sense of pleasure within the audience. Aesthetic Evaluation aims at evaluation/rating a painting or an artwork on the basis of various parameters like style, topic, emotional engagement etc. Aesthetic judgment of a piece of art cannot be done by a machine alone. The evaluation not only depends on the theme and style of the artist but also depends upon the likings, disliking, tastes and personal views of individuals evaluating the art. A piece of art which pleases some portions of a society need not necessarily please others. So we cannot leave behind the human inputs while determining the aesthetic value. In this project, we demonstrate a method to judge or evaluate the aesthetic value of a painting by combining the human inputs with several image features like texture, color, edges and scene.

Our approach in this project is to first classify the paintings into various categories and then judge them aesthetically based on some image features and a survey conducted among people of different age group and gender.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Motivation

In the present Digital Era with the advancements in technology and increase in number of smart phones and computers has eased image retrieval to large extent. Now a user can easily search for an image by giving a simple query on his smart phone device. The concept of aesthetic evaluation will help in retrieving the images based on personal taste and choice on the basis of aesthetic score.

The objective of theaesthetic assessment is to design methods which can automatically predict the perceived quality of a painting or an artwork. Such methods find great applications in the field of image retrieval. A system that can automatically evaluate painting aesthetics has many potential applications. Using image retrieval systems, similar images (from a content-based perspective) could be re-ranked using aesthetic properties. They could help a user to select the best pictures from his collection to make photo albums. Also, these models could be deployed directly in photo cameras to make real-time suggestions and give on the spot scores to the photographs clicked.

Knowing the aesthetic value of paintings will help in painting search and painting image management. Using the said concept we can develop a recommendation application which will show theuser a painting of required aesthetic quality. It will also be helpful in easing the process of searching a painting reducing both time and resources.Automatic evaluation using computer systems will give one more option/ guideline to designers and artists to evaluate and refine their ideas.

1.2 Problem statement

Classification of painting and artworks into different genres and then evaluating them on the basis of various image features like color, texture, lines etc. along with human input i.e. user ratings.

1.3 Contributions

The major contributions of this project are:

1. *Developing an Image Crawler: We developed an image crawler that could crawl the internet and download paintings of required category.*

2. *Classification of paintings into different categories:* We have used the convolutional neural network architecture and a SVM classifier to classify the digital images of paintings.

3. *Identification and extraction of image features:* After thorough research and discussions we found out that colors, texture and edges play an important role in deciding the aesthetic quality of a painting. We extracted color histograms, local binary patterns, ahistogram of oriented gradients and GIST features of all the paintings.

4. *Design a regression model:*We have trained a regression model for determining Aesthetic ratings of unrated paintings.

1.4Organization of the report

Chapter 2 details the literature review and background study of the algorithms and techniques used in this project.

Chapter 3 explains the implementation methodology of our project, flow charts, software and hardware platform used.

Chapter 4 discusses the experimental results obtained over several input painting images illustrating the algorithm performance

Chapter 5 gives the conclusion and future work that may ensue from this project.

CHAPTER 2:
BACKGROUND STUDY

2.1 LITERATURE REVIEW

Aesthetic Evaluation of images and photographs is a popular topic for research in Computer Vision field. There has been a lot of work done for assessing the aesthetic value of images but there are very few who have assessed the aesthetic value of Paintings & Artworks.

AVA dataset is one of the datasets widely used for assessing the aesthetic ratings. This large scale dataset containing more than 250,000 images was introduced by Murray et al. In their original work they formulated a binary classification problem and established the experimental settings. They computed Fisher Vector signatures from SIFT descriptors and trained an SVM which achieved maximum accuracy of 67%.

Lu et al. used convolutional neural networks on the AVA dataset and achieved classification accuracy between 60.25% and 71.2%. The architecture of convolutional network used by them contained 4 convolutional layers and 2 layers that were fully connected.

Bianco et al. used deep Convolutional Neural Network to predict image aesthetics. They fine tuned CNN architecture by casting the image aesthetic prediction as a regression problem and used AVA dataset.Their Experimental results show the robustness of the solution proposed, which outperforms the best solution in the state of the art by almost 17 % in terms (MRSSE).

Marchesotti et al. proposed to use generic image descriptors to assess aesthetic quality instead of hand-crafting features which would correlate with best photographic practices and achieved good results.

2.2 ALGORITHMS USED

2.2.1 GIST

GIST descriptor was first proposed in for scene recognition. It is based on the low dimensional representation of the scene that is called Spatial Envelope. They define the features that separate a scene from the rest. Those features that represent the dominant spatial structure of a scene are naturalness, openness, roughness, expansion and ruggedness. A multidimensional space is created to find out which scenes share membership in semantic categories such as street and highways by projecting shared memberships are projected closed together. The success of GIST in scene recognition supports that modeling a holistic representation of the scene is informative enough about scenes semantic categories.

Given an input image, a GIST descriptor is computed by

- Convolve the image with 32 Gabor filters at 4 scales, 8 orientations, producing 32 feature maps of the same size of the input image.
- Divide each feature map into 16 regions (by a 4x4 grid), and then average the feature values within each region.
- Concatenate the 16 averaged values of all 32 feature maps, resulting in a 16x32=512 GIST descriptor.
- Intuitively, GIST summarizes the gradient information (scales and orientations) for different parts of an image, which provides a rough description (the gist) of the scene.

Fig. 2.1GIST Descriptor for a Flower Painting

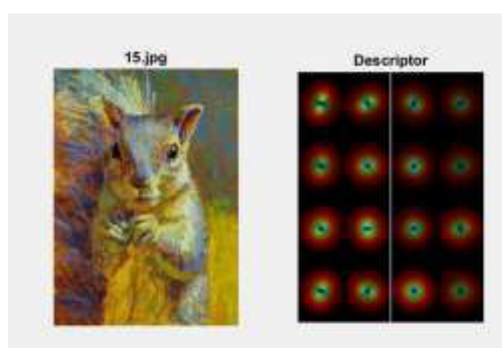


Fig. 2.2 GIST Descriptor for an Animal Painting

2.2.2 LBP (Local Binary Pattern)

It is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings.

Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. LBP is the particular case of the Texture Spectrum model proposed in 1990.[1][2] LBP was first described in 1994.[3][4] It has since been found to be a powerful feature for texture classification; it has further been determined that when LBP is combined with the Histogram of oriented gradients (HOG) descriptor, it improves the detection performance considerably on some datasets.[5] A comparison of several improvements of the original LBP in the field of background subtraction was made in 2015 by Silva et al.[6] A full survey of the different versions of LBP can be found in Bouwmans et al.[7]

The LBP feature vector, in its simplest form, is created in the following manner:

- Divide the examined window into cells (e.g. 16x16 pixels for each cell).
- For each pixel in a cell, compare the pixel to each of its 8 neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle, i.e. clockwise or counter-clockwise.
- Where the center pixel's value is greater than the neighbor's value, write "0". Otherwise, write "1". This gives an 8-digit binary number (which is usually converted to decimal for convenience).
- Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center). This histogram can be seen as a 256-dimensional feature vector.
- Optionally normalize the histogram.

- Concatenate (normalized) histograms of all cells. This gives a feature vector for the entire window.

The feature vector can now be processed using the Support vector machine, extreme learning machines, or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis.

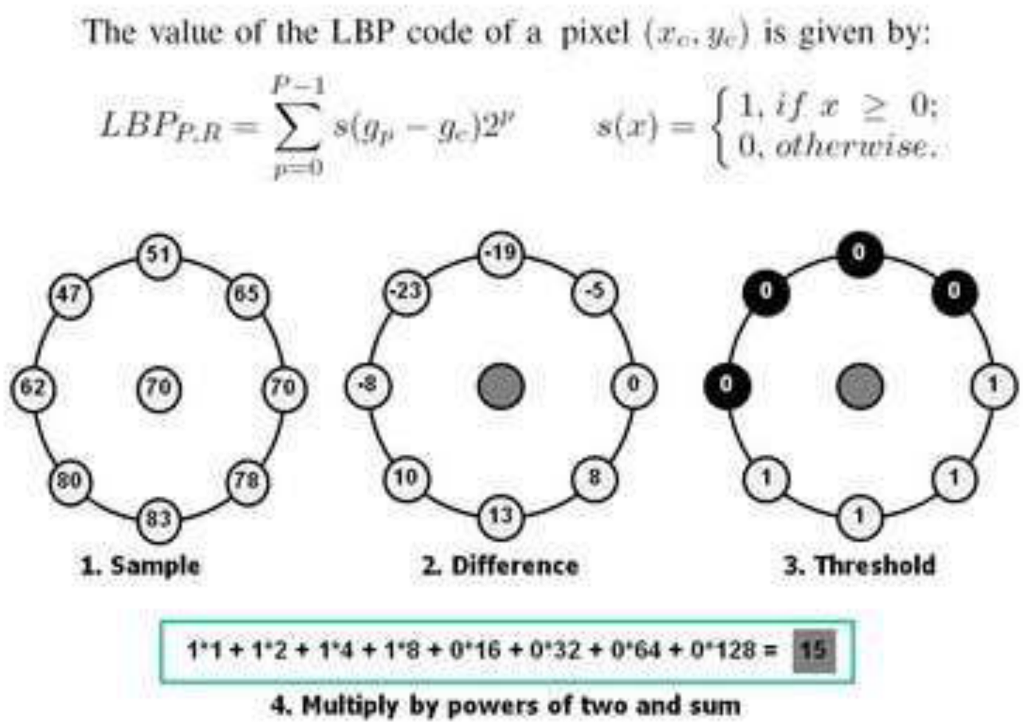


Fig. 2.3 LBP

The following notation is used for the LBP operator: $LBP_{P,R}^{u2}$. The subscript represents using the operator in a (P,R) neighborhood. Superscript u2 stands for using only uniform patterns and labeling all remaining patterns with a single label. After the LBP labeled image $f_l(x,y)$ has been obtained, the LBP histogram can be defined as

$$H_i = \sum_{x,y} I\{f_l(x,y) = i\}, i = 0, \ldots, n-1,$$

in which n is the number of different labels produced by the LBP operator, and I{A} is 1 if A is true and 0 if A is false.

When the image patches whose histograms are to be compared have different sizes, the histograms must be normalized to get a coherent description:

$$N_i = \frac{H_i}{\sum_{j=0}^{n-1} H_j}.$$

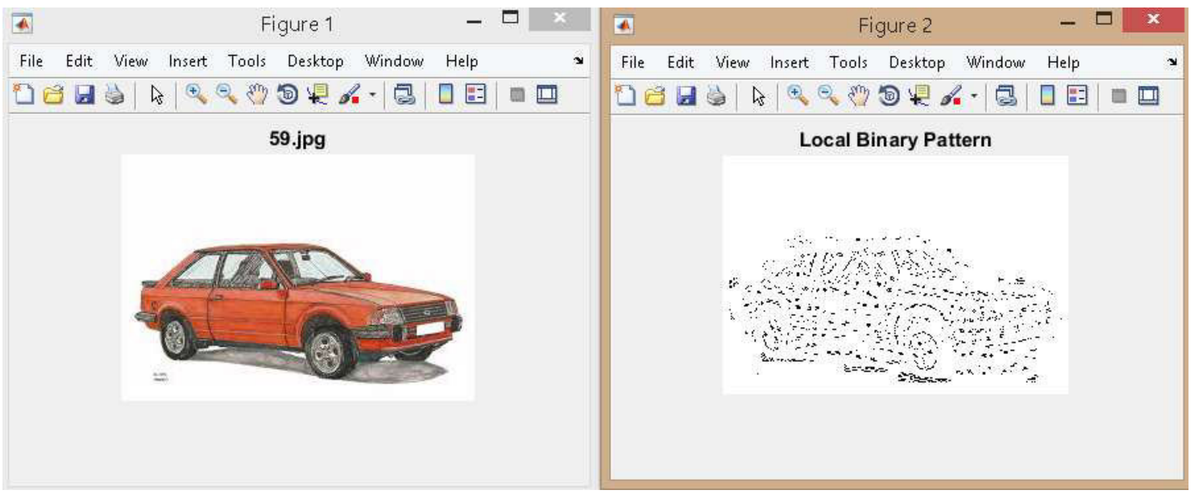


Fig. 2.4 Local Binary Pattern of a Car

2.2.3 HOG (Histogram of Oriented Gradients)

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Robert K. McConnell of Wayland Research Inc. first described the concepts behind HOG without using the term HOG in a patent application in 1986.[1] In 1994 the concepts were used by Mitsubishi Electric Research Laboratories.[2] However, usage only became widespread in 2005 when Navneet Dalal and Bill Triggs, researchers for the French National Institute for Research in Computer Science and Automation (INRIA), presented their supplementary work on HOG descriptors at the Conference on Computer Vision and Pattern Recognition (CVPR). In this work they focused on pedestrian detection in static images, although since then they expanded their tests to include human detection in videos, as well as to a variety of common animals and vehicles in static imagery.

The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing.

The HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. Moreover, as Dalal and Triggs discovered, coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permit the individual body movement of pedestrians to be ignored so long as they maintain a roughly upright position.

Gradient computation

The first step of calculation in many feature detectors in image preprocessing is to ensure normalized color and gamma values. As Dalal and Triggs point out, however, this step can be omitted in HOG descriptor computation, as the ensuing descriptor normalization essentially achieves the same result. Image pre-processing thus provides a little impact on performance. Instead, the first step of calculation is the computation of the gradient values. The most common method is to apply the 1-D centered, point discrete derivative mask in one or both of the horizontal and vertical directions. Specifically, this method requires filtering the color or intensity data of the image with the following filter kernels:

$$[-1, 0, 1] \text{ and } [-1, 0, 1]^\top .$$

Orientation binning

The second step of calculation is creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The cells themselves can either be rectangular or radial in shape, and the histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is “unsigned” or “signed”. Dalal and Triggs found that unsigned gradients used in conjunction with 9 histogram channels performed best in their human detection experiments. As for the vote weight, pixel contribution can either be the gradient magnitude itself, or some function of the magnitude. In tests, the gradient magnitude itself generally produces the best results.

Fig. 2.5 HOG Features plotted on paintings.

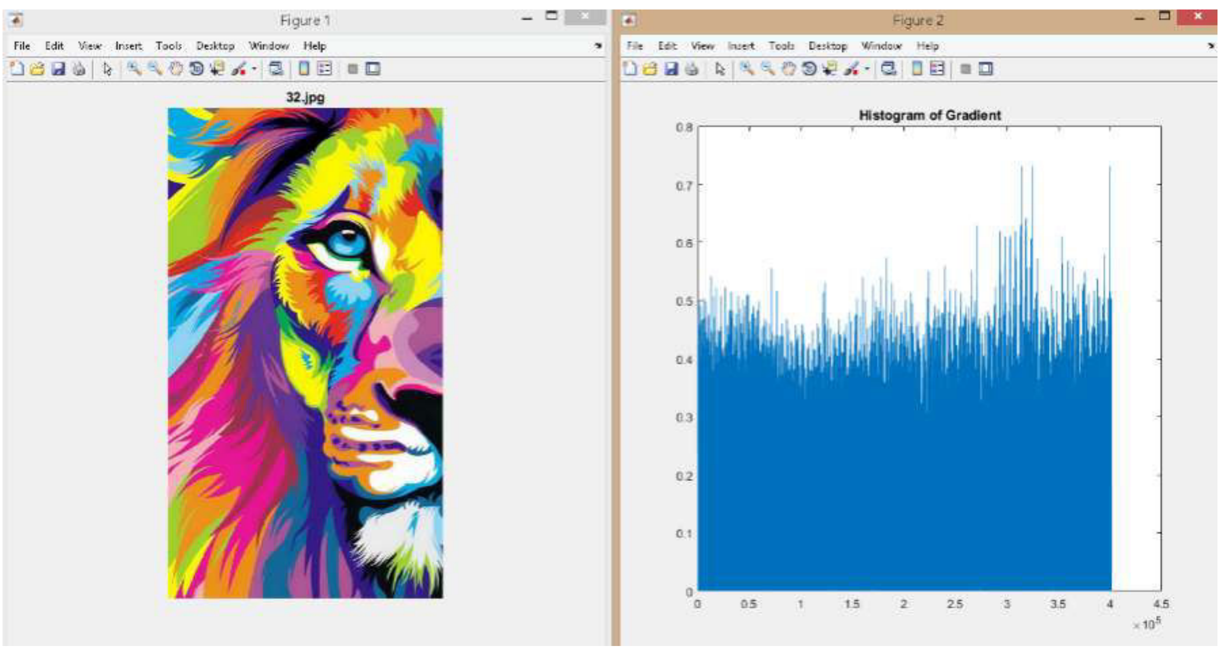


Fig. 2.6 Histogram of Gradient.

2.2.4 Colour Histogram

In image processing and photography, a color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image's color space, the set of all possible colors.

The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. For monochromatic images, the term intensity histogram may be used instead. For multi-spectral images, where each pixel is represented by an arbitrary number of measurements (for example, beyond the three measurements in RGB), the color histogram is N-dimensional, with N being the number of measurements taken. Each measurement has its own wavelength range of the light spectrum, some of which may be outside the visible spectrum.

If the set of possible color values is sufficiently small, each of those colors may be placed on a range by itself; then the histogram is merely the count of pixels that have each possible color. Most often, the space is divided into an appropriate number of ranges, often arranged as a regular grid, each containing many similar color values. The color histogram may also be represented and displayed as a smooth function defined over the color space that approximates the pixel counts.

Like other kinds of histograms, the color histogram is a statistic that can be viewed as an approximation of an underlying continuous distribution of colors values.A color histogram focuses only on the proportion of the number of different types of colors, regardless of the spatial location of the colors. The values of a color histogram are from statistics. They show the statistical distribution of colors and the essential tone of an image.In general, as the color distributions of the foreground and background in an image are different, there might be a bimodal distribution in the histogram.

For the luminance histogram alone, there is no perfect histogram and in general, the histogram can tell whether it is over exposure or not, but there are times when you might think the image is over exposed by viewing the histogram; however, in reality it is not.

The main drawback of histograms for classification is that the representation is dependent of the color of the object being studied, ignoring its shape and texture. Color histograms can potentially be identical for two images with different object content which happens to share color information. Conversely, without spatial or shape information, similar objects of different color may be indistinguishable based solely on color histogram comparisons. There is no way to distinguish a red and white cup from a red and white plate. Put another way, histogram-based algorithms have no concept of a generic 'cup', and a model of a red and white cup is no use when given an otherwise identical blue and white cup. Another problem is that color histograms have high sensitivity to noisy interference such as lighting intensity changes and quantization errors. High dimensionality (bins) color histograms are also another issue. Some color histogram feature spaces often occupy more than one hundred dimensions.

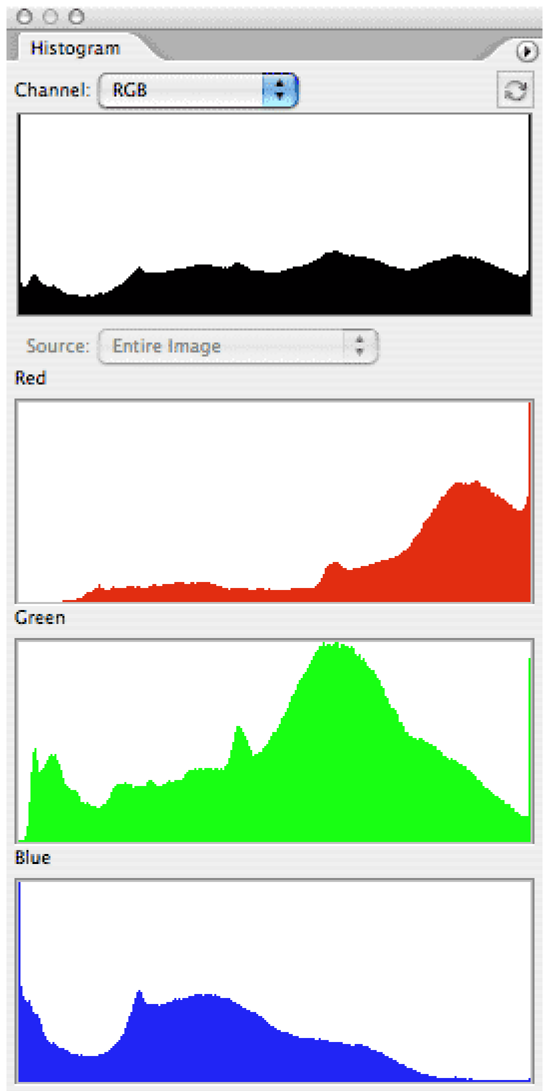


Fig. An example of Colour Histogram.

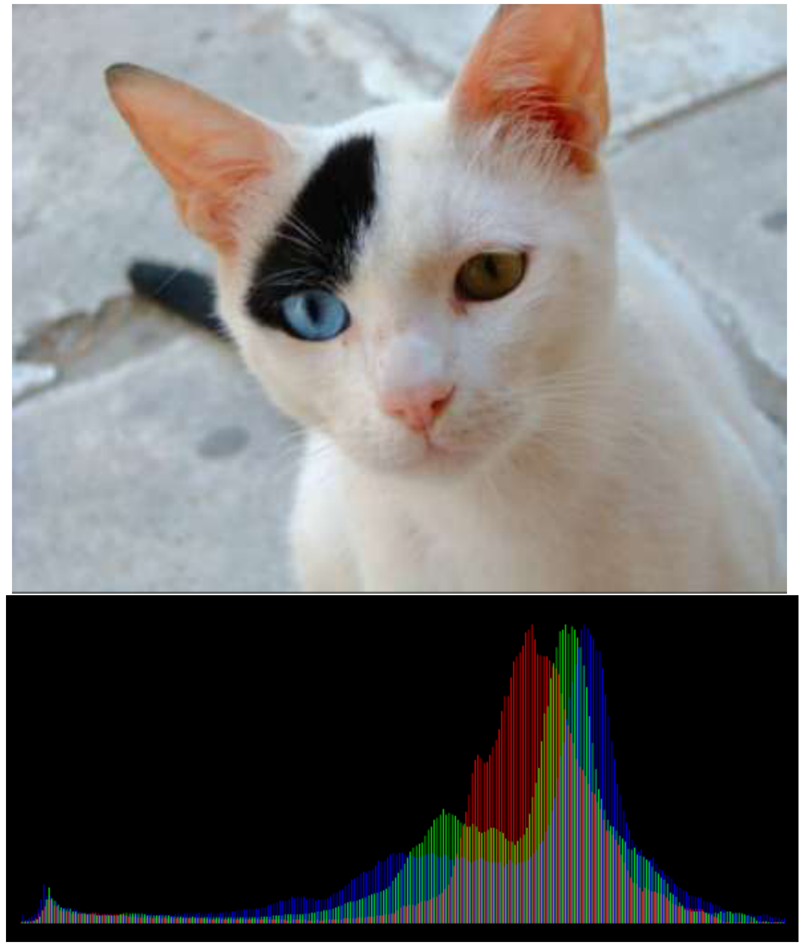


Fig. 2.7 An image of a cat and its Colour Histogram.

2.2.5Convolutional Neural Network

It is a class of deep, feed-forward artificial neural network that has successfully been applied to analyzing visual imagery. CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing, They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

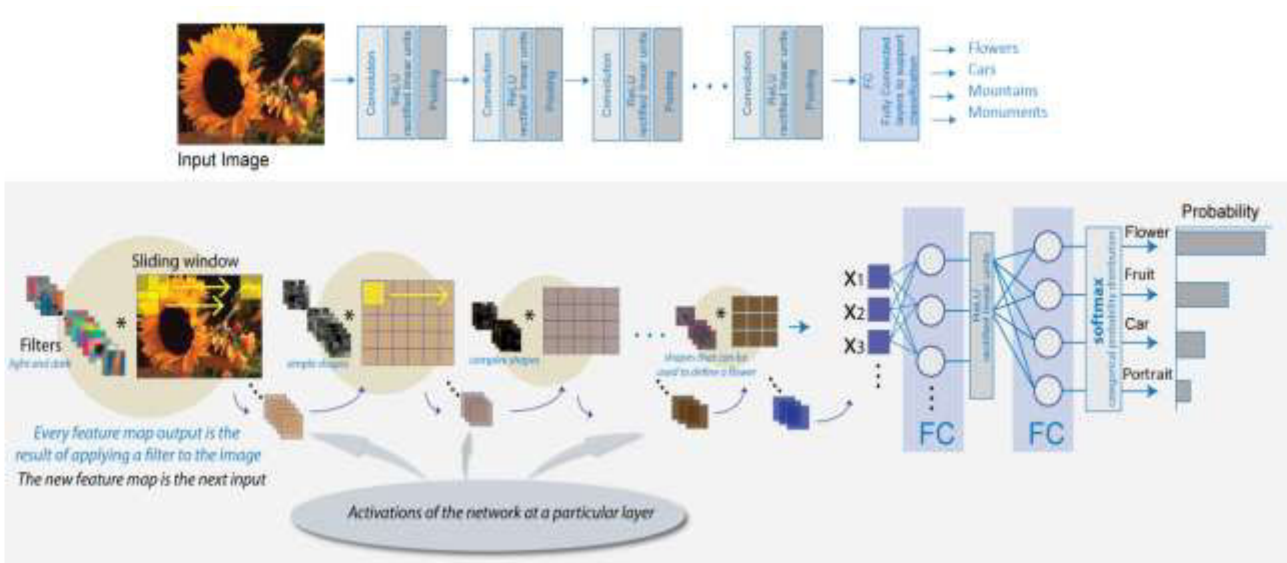


Fig. 2.7 ConvNet Architecture

Convolutional networks were inspired by biological processes in which the connectivity pattern between neurons is inspired by the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage.

They have applications in image and video recognition, recommender systems and natural language processing.

AlexNet: Convolutional Neural Network

AlexNet (designed by Krizhevsky et al. [11] is one of the deep ConvNets designed to deal with complex scene classification task on Imagenet data. The task is to classify the given input into one of the 1000 classes.The main differences between LeNet and AlexNet are in the i) Number of processing layers and number of trainable parameters: AlexNet has 5 convolutional layers, 3 sub sampling layers, 3 fully connected layers and its total trainable parameters are in crores whereas LeNet has 2 convolutional layers, 2 Sub sampling layers and 3 fully connected layers and its total trainable parameters are in thousands. ii) The non linearity used in the feature extractor module of the AlexNet is ReLU whereas LeNet has logistic sigmoid. iii) AlexNet uses dropout where as no such concept is used in LeNet. A fully trained AlexNet on ImageNet data set can not only be used to classify Imagenet data set but it can also be used without the output layer to extract features from samples of any other data set.

The arrangement and configuration of all the layers of AlexNet is shown in Fig. 3. As can be seen in the figure, it uses a different activation function called Rectified Linear Unit (ReLU) (described in Section 4a) after every Convolutional layers and it also has a new type of processing called dropout after fully connected layers 1 and 2, the same is described in Page 5. The arrangement of Convolutional and pooling layers reduces the number of input features from 154587 to 1024 before sending them to the fully connected layers. Like LeNet, the Convolutional layers of AlexNet also takes inputs from a subset of feature maps generated by the previous layer. The detailed mapping is described in the work of Krizhevsky et al.

Yellow square indicates the input feature map. Green indicates the convolutional layer, Orange indicates the max pooling layer and Blue indicates the fully connected layer. Arrows show the direction of flow of data. Dropout 0.5 indicates that a dropout layer exist between two fully connected layers with a retain probability of 0.5. Output indicates the number of feature maps a layer generates and its dimension is specified by Size. RFS indicates Receptive Field Size of the layer. Strides mention the horizontal and vertical jump of the receptive field. The complete architecture has 11 processing layers and more than 2 crore trainable weights.

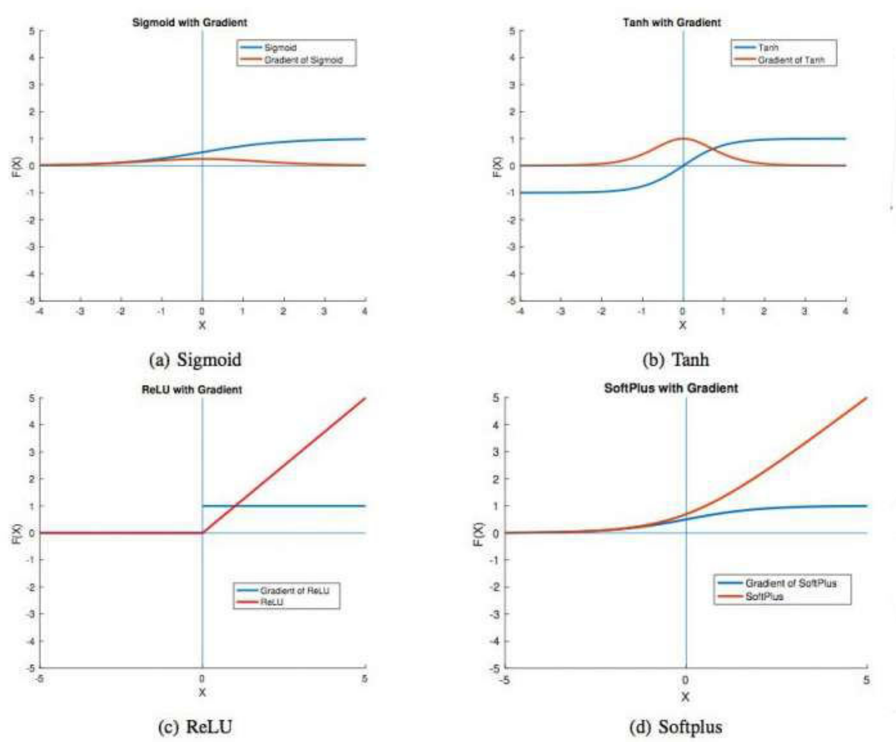
ReLU: Rectified Linear Unit

ReLU is a function first introduced by Hahnloser et al. [12]. It was then stated by Nair et al. [13] that ReLU is an effective activation function for use in neural network as well. ReLU function is given by:

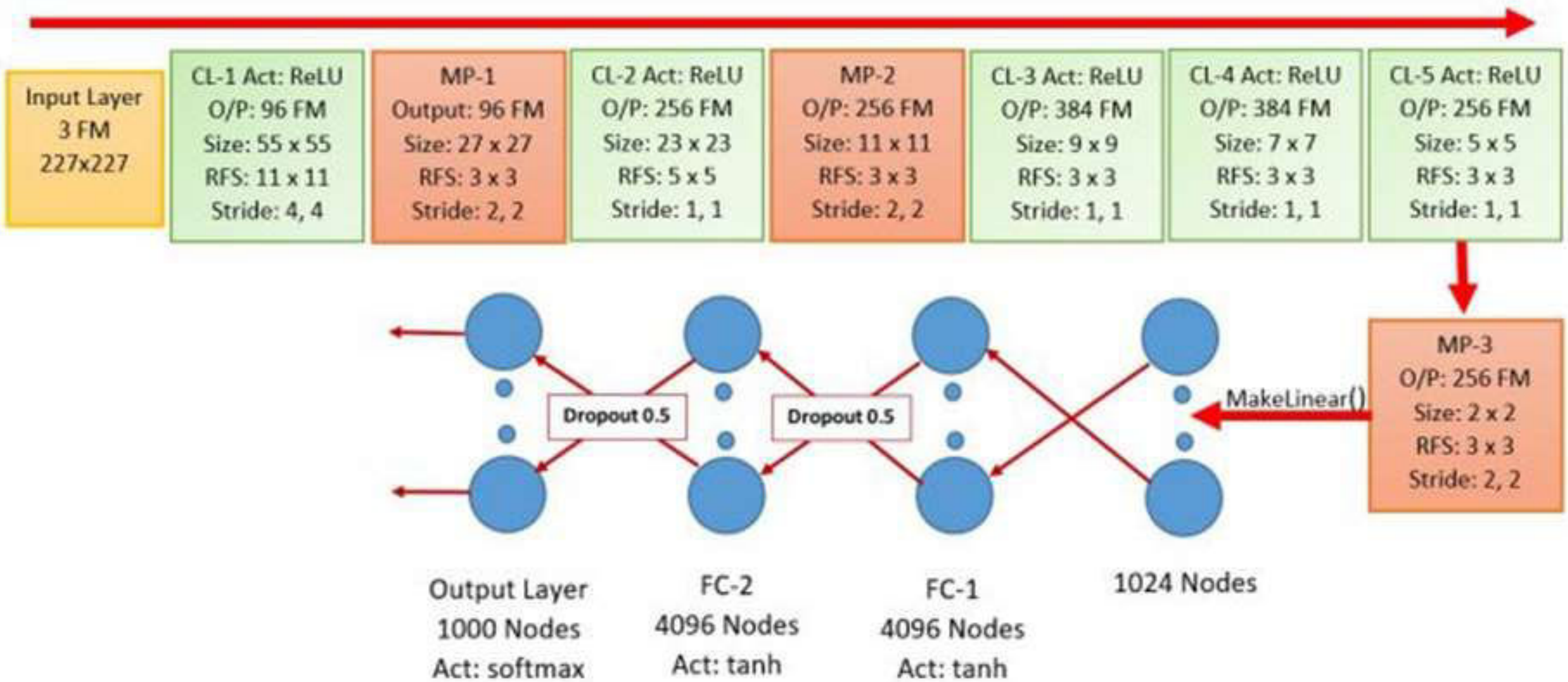
$$f(x) = \max(0,x)$$

The plot of ReLU along with the plot of its derivative is shown in Fig. 4c and the corresponding expressions are given in Table II. ReLU is applied after the Convolutional layer, to induce sparsity in the features and to solve the problem of vanishing gradient. ReLU is not differentiable at 0 and this creates problem during back propagation which requires derivative at all points to be defined. SoftPlus which is shown in Fig. 4d is a smooth approximation to ReLU which was proposed by Dugas et al. [14] and is given by: f(x) = log(1+ex)

Softplus is differentiable at 0 but it is still not used popularly due to computations involved in evaluating the exponent and logarithm functions. Instead ReLU is used by considering the derivative at 0 to be 0 or some small values ε.



(a) Sigmoid (b) Tanh (c) ReLU (d) Softplus



AlexNet Architecture

ReLU has a disadvantage that the network using it suffers from the dying ReLU problem. The problem comes when a node generates a negative output. In such cases ReLU generates a derivative of 0, during backward pass and because of this the weights attached to the node behind ReLU are not trained. As they are not trained there is not much significant change in the value of the node in consecutive forward pass and the node does not get a chance to recover from the negative value, thus the node becomes potentially dead. When most of the nodes of a layer generate negative values, lesser is the training of the weights behind that layer. In an extreme case, if all the outputs from a layer become negative then no further training will be done for the weights behind that layer in consecutive training iterations.

The two main reasons for using ReLU in Convolutional layers are

1. Faster convergence due to non existence of vanishing gradient problem. The gradient is made up of a product of derivatives of many activation functions which comes in the way from the node whose weights are to be updated to the output layer. In case of hyperbolic tangent these derivatives become very small when the output of the nodes goes in the saturating zone of the hyperbolic tangent function. As a result the product of such small derivatives which forms part of the gradient, becomes extremely small causing the gradient to vanish and this phenomena is also called as the vanishing gradient problem. Convolutional layer suffers more from these as they are many layers away from the output layer. The same does not happen with ReLU activation function (used in Convolutional layers), as it does not have a saturating zone for positive inputs.

2. **Inducing sparsity in features**: Convolutional layer extract feature unlike fully connected layers. Feature extraction requires sparsity in the input feature maps and it should set to 0 as many features as possible. The features preserved with non zero values are those that are discriminative in nature. It can be observed that all the features which are required to identify the rectangle are preserved and all others are set to 0 (represented in black color). The sparsity does not come into effect with other activation functions as they can generate small values instead of zeros. The sparsity in the features helps in speeding up the computation process by removing the undesired features.

2.2.6 Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyperplane that differentiates the two classes very well (look at the below snapshot).

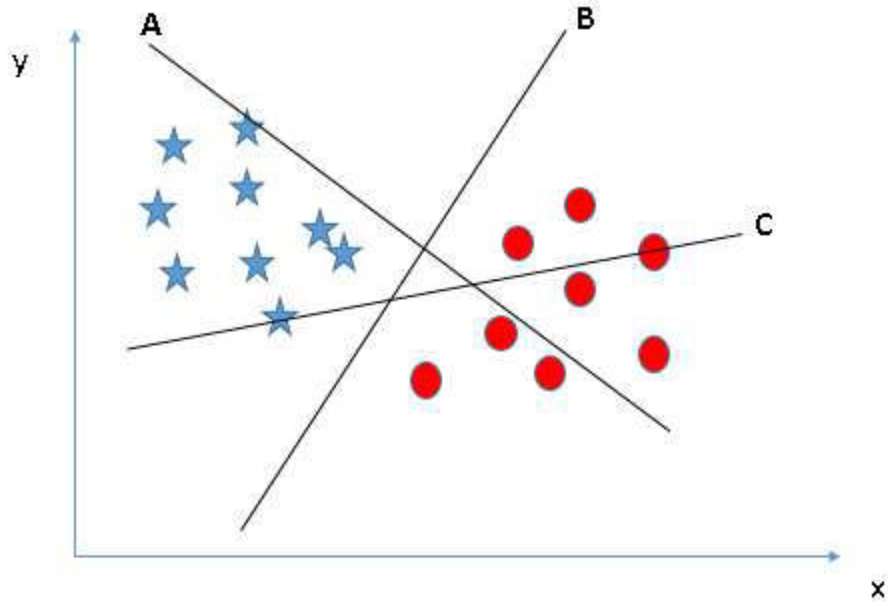


Fig. 2.8 SVM Classification

Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyperplane/ line).

The SVM algorithm is implemented in practice using a kernel.

The learning of the hyperplane in linear SVM is done by transforming the problem using some linear algebra, which is out of the scope of this introduction to SVM.A powerful insight is that the linear SVM can be rephrased using the inner product of any two given observations, rather than the observations themselves. The inner product between two vectors is the sum of the multiplication of each pair of input values.

For example, the inner product of the vectors [2, 3] and [5, 6] is 2*5 + 3*6 or 28.

The equation for making a prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows:

$$f(x) = B0 + \text{sum}(ai * (x,xi))$$

This is an equation that involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai (for each input) must be estimated from the training data by the learning algorithm.

Linear Kernel SVM

The dot-product is called the kernel and can be re-written as:

$$K(x, xi) = \text{sum}(x * xi)$$

The kernel defines the similarity or a distance measure between new data and the support vectors. The dot product is the similarity measure used for linear SVM or a linear kernel because the distance is a linear combination of the inputs.

Other kernels can be used that transform the input space into higher dimensions such as a Polynomial Kernel and a Radial Kernel. This is called the Kernel Trick.

It is desirable to use more complex kernels as it allows lines to separate the classes that are curved or even more complex. This in turn can lead to more accurate classifiers.

Polynomial Kernel SVM

Instead of the dot-product, we can use a polynomial kernel, for example:

$$K(x,xi) = 1 + \text{sum}(x * xi)^\wedge d$$

Where the degree of the polynomial must be specified by hand to the learning algorithm. When d=1 this is the same as the linear kernel. The polynomial kernel allows for curved lines in the input space.

Radial Kernel SVM

Finally, we can also have a more complex radial kernel. For example:

$$K(x,xi) = exp(-gamma * sum((x – xi^2))$$

Where gamma is a parameter that must be specified to the learning algorithm. A good default value for gamma is 0.1, where gamma is often 0 < gamma < 1. The radial kernel is very local and can create complex regions within the feature space, like closed polygons in two-dimensional space.

CHAPTER 3:
IMPLEMENTATION METHODOLOGY

3.1 FLOW-CHARTS

A flowchart is a type of diagram that represents an algorithm, workflow or process. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields.

The following flowcharts describes our basic approach of training a regression model where we extract image features of all paintings of our dataset. Then the extracted features are fed into a regression model as predictor variables and the user ratings on the paintings are given as response variable.

For predicting an aesthetic score of an image, we extracted the image features and passed them to the trained model. The model then predicts out the aesthetic score of the test image.

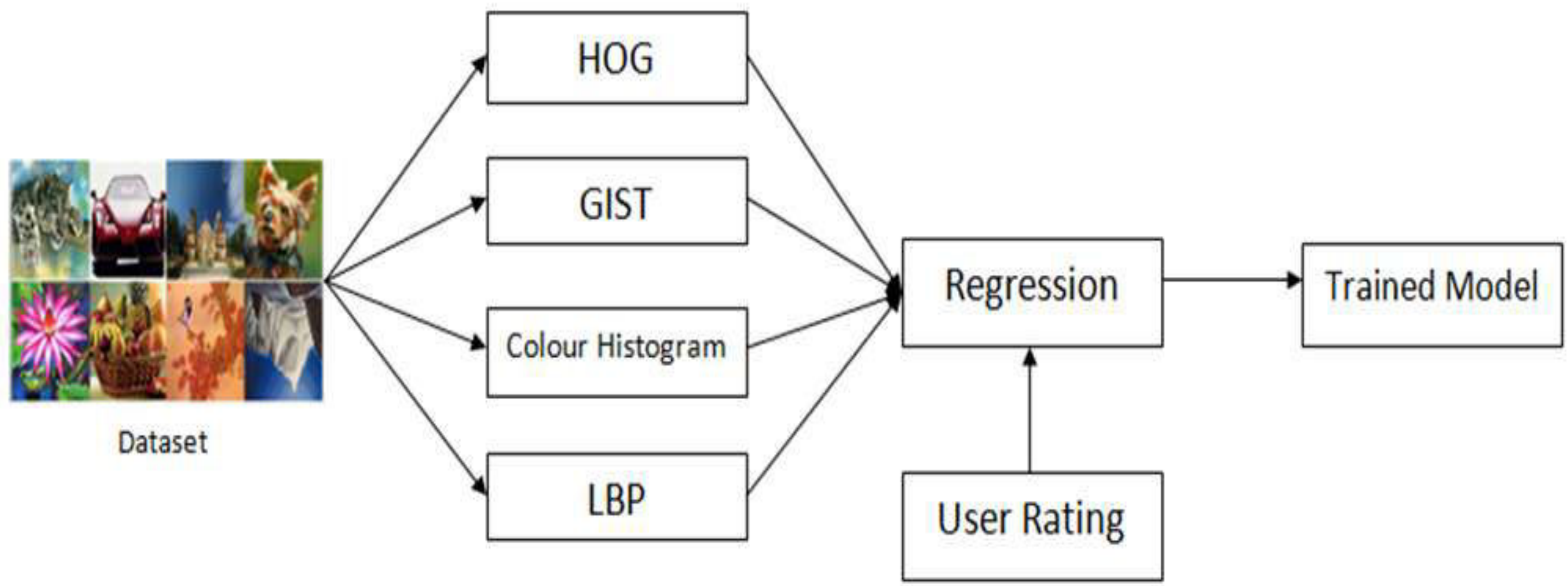


Fig. 3.1 Training a Regression Model

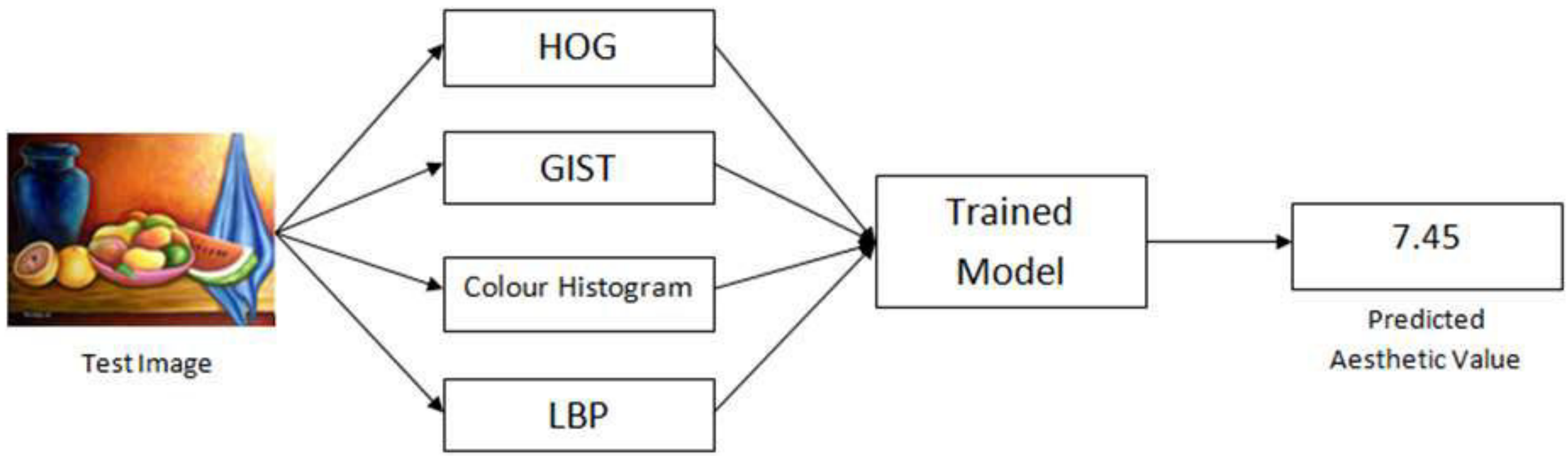


Fig. 3.2 Predicting Aesthetic Value using the trained model

3.2 Languages Used

1. MATLAB (for Feature Extraction/Classification/Evaluation)
2. Python (for developing a crawler to download images from the Internet)

3.3 Dataset

We downloaded various paintings and artworks of 7categories viz. Animal paintings, Fruit Paintings, Portrait Paintings, Car Paintings, Monument Paintings, Mountain Paintings and Flower Paintings. We developed an image crawler in Python to download images from the Internet and downloaded 175 paintings of each category, making the image count of the dataset to 1225, which was manually cross-checked. For evaluation purpose, we collected ratings on each of the 1225 paintings on a scale of 1 to 10. A rating close to 1 means the image is not aesthetically good whereas rating close to 10 implies the image to be beautiful and aesthetically good.

3.4 Machine Used

Processor: Intel Core i5

GPU: Nvidia 830M

Memory: 4.00 GB

System Type: 64-bit Windows 8.1Operating System

CHAPTER 4: EXPERIMENTAL RESULT

4.1 Classification Of Paintings

The following table determines the accuracy achieved using various techniques. It is evident from the data that CNN works the best giving an accuracy of 92.85%.

Algorithm	Bag of Features	HOG	GIST	CNN
Accuracy	54.5%	59.6%	61.2%	92.85%

Table. 4.1 Accuracy achieved using different algorithms/techniques

Classification Predictions made by our model:

```
label4 =

     Animal_Painting
```

Fig 4.1 Sample Run 1: Correctly predicted an animal

	animal	car	flower	fruits	monument	mountain	portrait
animal	**0.9000**	0	0.0333	0.0333	0	0.0333	0
car	0	**1.0000**	0	0	0	0	0
flower	0.1333	0	**0.7667**	0.0667	0	0.0333	0
fruits	0	0	0.0333	**0.9000**	0	0.0333	0.0333
monument	0	0	0	0	**0.9667**	0.0333	0
mountain	0	0	0	0	0.0333	**0.9667**	0
portrait	0	0	0	0	0	0	**1.0000**

Accuracy :
92.8571

Table 4.2 Confusion Matrix obtained using Convolutional Neural Network

1

4.2 Aesthetic Evaluation Of Paintings

We extracted the following features: GIST, Histogram of Oriented Gradients, Local Binary Pattern, Colour Histograms. Then we trained a regression model with above-mentioned predictor variables against the user ratings collected from persons of different age group and gender.

Since Aesthetic Value of an image is not a fixed value, we used the following method to determine the accuracy of our prediction:

If **abs (predicted_value_rating - average_user_rating) <=1**, the prediction is said to be correct. Using the above-said methodology we achieved **64.7%** accuracy.

Aesthetic Evaluation Predictions made by our model:

Fig. 4.3 Sample Run 1: Prediction = 8.4275

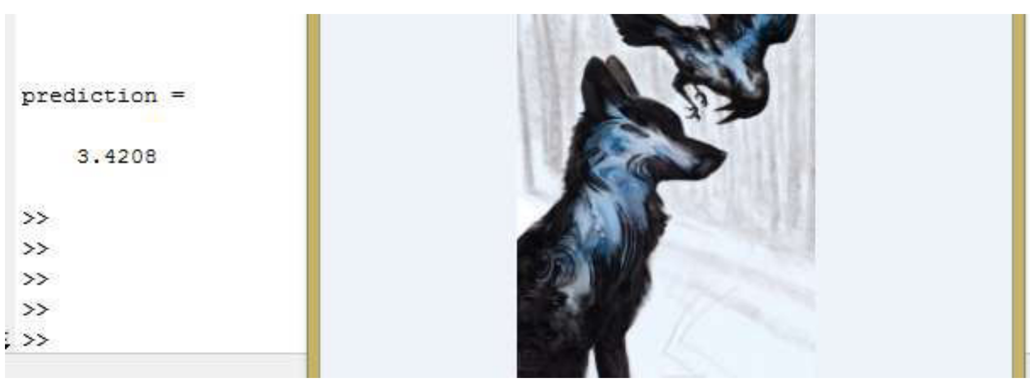


Fig. 4.4 Sample Run 2: Prediction = 3.4208

CHAPTER 5:
CONCLUSION AND FUTURE WORK

The experiments and data analysis in this project investigated a machine learning solution to aesthetic evaluation of paintings and artworks. The classification is done using CNN and SVM classifier performed with an accuracy of more than 92% for over thousand paintings of seven categories. Evaluation of a painting is a complex task which sometimes depends on user's taste and understanding. We achieved 64% accuracy by using Image Features like HOG, LBP, GIST etc combined with inputs from 100 persons on each painting.

From the accuracy achievedit can be concluded that using image features and machine learning approach we can evaluate the aesthetic value of paintings. It is highly possible to improve the performance of the model by using more featured and taking ratings from people of different domains.

After improving the performance of the model it may be helpful in optimizing the search engines to show relevant images and paintings to the users. The search engine will show the results based on the aesthetic value input by the user. It would also be helpful in enhanced management of paintings. Art galleries and museums would be highly benefited by the aesthetic evaluation as it will help in attracting the audience.

More complex implementation can enable robots to automatically paint as per the requirement of the users or evaluate and provide guidelines to improve and enhance already existing artworks.

REFERENCES

[1] A. Oliva and A. Torralba. "Modeling the shape of the scene: A holistic representation of the spatial envelope." Int. J. Comput. Vision, 42(3):145–175, May 2001.

[2] P.Veerina."Learning Good Taste: Classifying Aesthetic Images"

[3] A. Joy, Sreekumar K. "Aesthetic Quality Classification of Photographs: A Literature Survey"

[4] N. Murray, L. Marchesotti, F. Perronnin."AVA: A Large-Scale Database for Aesthetic Visual Analysis".

[5] A. Krizhevsky, I. Sutskever, G. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks"

[6] S. Khan and D.Vogel Evaluating. "Visual Aesthetics in Photographic Portraiture". Computational Aesthetics in Graphics, Visualization, and Imaging (2012), pp. 1–8

[7] Wei-Ta Chu. "Aesthetic Visual Quality Assessment of Paintings"

[8] L. Marchesotti, F. Perronnin, D. Larlus, G. Csurka ."Assessing the aesthetic quality of photographs using generic image descriptors"