TOWARDS AUTOMATED CLASSIFICATION OF FINE-ART PAINTING STYLE: A COMPARATIVE STUDY

By

RAVNEET SINGH ARORA

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Computer Science

written under the direction of

Dr. Ahmed Elgammal

and approved by

____

____

____

New Brunswick, New Jersey

October, 2012

ABSTRACT OF THE THESIS

Towards Automated Classification of Fine-art Painting Style: a Comparative Study

by RAVNEET SINGH ARORA

Thesis Director:
Dr. Ahmed Elgammal

This thesis presents a comparative study of different classification methodologies for the task of fine-art genre classification. The problem of painting classification involves classifying new unknown paintings among different art genres. Two-level comparative study is performed for this classification problem. The first level reviews the performance of discriminative vs. generative models while the second level touches the features aspect of the paintings and compares Semantic-level features vs low-level and intermediate-level features present in the painting. Three models are studied and compared, namely - 1) A Discriminative model using a Bag-of-Words (BoW) approach; 2) A Generative model using BoW; 3) Discriminative model using Semantic-level features. Various experiments and techniques like Bag of Words model, Topic models and Classeme features are employed to get insights into potential of these automatic classification techniques for painting styles.

Acknowledgements

I gratefully acknowledge the guidance and generous support of my supervisor, Professor Ahmed Elgammal, throughout my masters. His advice and support helped me overcome various obstacles during this period. I am thankful for his contribution to this thesis as well as all my other studies.

I would also like to thank my fellow researchers at CBIM. Discussions and meetings we used to have proved really helpful for my research and thesis. I really enjoyed the flow of ideas we all shared during those meetings.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Objective

Paintings belong to a special class of images that contains intrinsic properties that haven't been studied in too much detail till now. This thesis focuses on the problem of paintings style genre classification. The primary goal is to study and perform a comparative evaluation of different methodologies for the automatic task of painting style classification.

The problem of painting style classification can be stated as – **Given a set of paintings for each painting style, predict the style of an unknown painting**. A lot of work has been done so far on the problem of image category recognition. But the problem of painting classification proves quite different than that of image category classification. Paintings are differentiated not only by contents but also by style applied by a particular painter or school of painting or by the age when they were painted. This also makes painting classification problem much more challenging than the ordinary image category recognition problem.

In this thesis we will approach the problem of painting style classification from a supervised learning perspective. Two level comparative study will be conducted for this classification problem. The first level reviews the performance of discriminative vs. generative models while the second level touches the feature aspects of the paintings and compares semantic-level features vs. low-level and intermediate-level features present in the painting.

For experimental purposes seven fine-art genre categories will be used, namely - *Renaissance, Baroque, Impressionism, Cubism, Abstract, Expressionism, and Popart*. **Figure 1** displays some

Figure 1. Painting Styles used

of the samples for each category. Various different set of comparative experiments focused on evaluation of classification accuracy for each methodology will be performed.

We will evaluate three different methodologies namely –

1. Discriminative model using a Bag-of-Words (BoW) approach
2. Generative model using BoW
3. Discriminative model using Semantic-level features

Figure 2 displays the organization of our comparative study

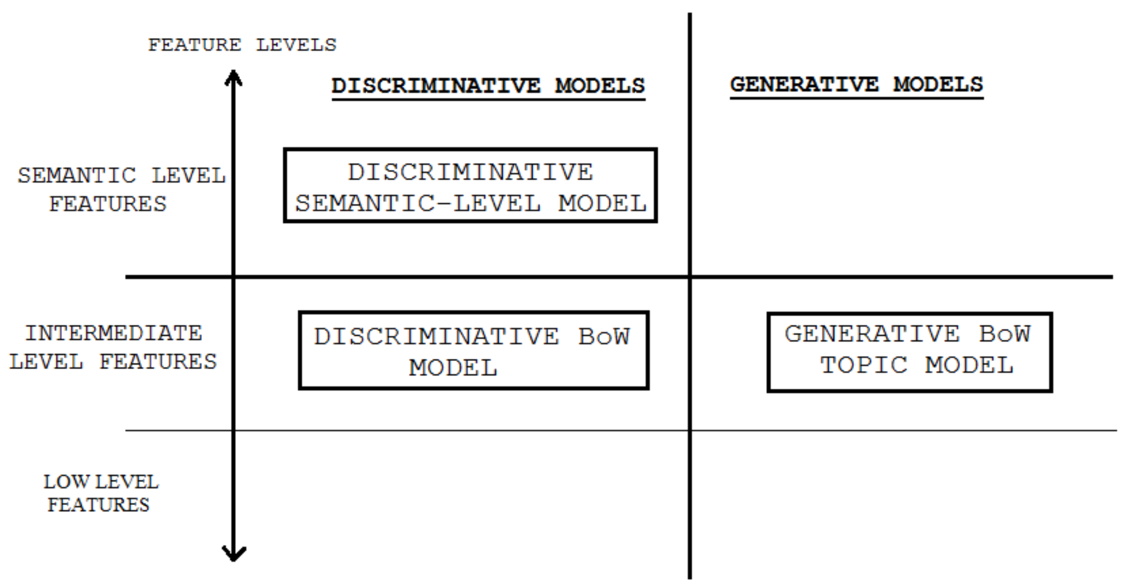

Figure 2. Classification Methodologies

As can be observed from **Figure 2**, these models differ from each other not only in terms of type of models used but also by the level of features the model works on. The Discriminative Semantic-Level model applies a discriminative machine learning model upon features capturing semantic information present in a painting, while Discriminative and Generative BoW models employs discriminative and generative machine learning models respectively on the Intermediate level features represented using BoW model.

A generative model has the property that it specifies a joint probability distribution over observed samples and their labels. In other words, a generative classifier learns a model of joint probability distribution p(x,y), where ‘x’ denotes the observed samples and ‘y’ are the labels. Bayes rule can be applied to predict the label ‘y’ for a given new sample ‘x’, which is determined by the probability distribution p(y|x). Since generative models calculates the distribution p(x|y) as an intermediate step, these can be used to generate random instances ‘x’ conditioned on target labels ‘y’.

A discriminative model on the other hand tries to estimate the distribution p(y|x) directly from the training data. Thus a discriminative model bypasses the calculation of joint probability distribution p(x,y) and avoids the use of Bayes rule. Andrew Y. Ng et al. [54] provides a comprehensive comparison of both models.

It is also very important to make distinction between Low, Intermediate and Semantic-level features at this stage. Low-level features capture directly the formal characteristics of paintings such as color, texture, edges, light etc. The average intensity of all the pixels, color histogram representing color composition of paintings and number of edges are few low-level features that captures the formal elements: light, color and edges respectively.

Intermediate-level features apply local-level descriptors like SIFT and CSIFT on various regions of an image. Local level descriptors instead of summarizing the whole image, represents localized regions of an image. A Bag of Words model is applied to generate an intermediate representation of the image. A Bag of Words model first creates fixed number of clusters from the localized regions of all the images and further represents each image by the histogram capturing the frequency of the clusters in that image. Such clustering of local features from all the images provide an intermediate-level representation.

Semantic-level features capture the semantic content classes such as water, sand, cars etc. present in an image. Thus, such frequency of semantic classes can help us in rank images according to their semantic similarity. A feature vector where each element denotes the probability of existence of a semantic class is an example of semantic feature.

It is worth noting that, instead of using low-level features like color, light, shades and texture our study is focused on intermediate-level features (BoW features) and semantic-level features.

We hypothesize the following claims – 1) Semantic level information hidden in a painting can be very well utilized for the task of classification and 2) Generative models like Topic models are very much capable of capturing the **thematic structure** of a painting**.** It is easy to visualize a topic or theme in the case of documents. For documents, a topic can be a collection of particular set of words. For example, a science topic is characterized by the collection of words like atom, electrons, protons etc. For images represented by a Bag of Word model, each word is represented by the local level descriptor used to describe the image. Thus a collection of particular set of such similar regions can constitute a topic. For example, collection of regions representing mainly straight edges can constitute the topic trees. Similarly, set of regions having high concentration of blue color can form up a theme related to sky or water. However, in images, it is much harder to relate topics to actual semantic concepts.

1.2 Motivation

In the last decade there have been impressive advances in developing computer vision algorithms for different object recognition-related problems including: instance recognition, categorization, scene recognition, pose estimation, etc. When we look into an image we not only recognize object categories, and scene category, we can also infer various cultural and historical aspects. For example, when we look at a fine-art paining, an expert or even an average person can infer information about the genre of that paining (e.g. Baroque vs. impressionism) or even can guess the artist who painted it. This is an impressive ability of human perception for learning complex visual concept.

Besides the scientific merit of the problem from the perception point of view, there are various application motivations. Paintings can provide a lot of rich heritage information about particular time-era it belongs to. With the increasing sizes of digitized art databases on the Internet comes the daunting task of organization and retrieval of paintings. There are millions of paintings present on internet. To manage properly the databases of these paintings, it becomes very necessary to classify paintings into different categories and sub-categories. This classification structure can be utilized as an index and thus can really improve the speed of retrieval process. Manual classification of such a large collection of images can prove to be very time consuming and would require a lot of effort. Thus it becomes very important to come up with an efficient and accurate classifier for paintings, which can be easily applied to such a large collection of paintings.

Automatic classifier for paintings can also act very useful in cases where identity of the paintings is unknown. It provides a way of getting insights into the origins of an unknown painting. Machine learning models can prove really useful in tasks of inferring parameters of an unidentified painting based on the already existing database of paintings. Once the parameters of a new painting are known it can be easily indexed into the already existing database of paintings.

Also the applications of paintings classification are not only restricted to personal computers. With the explosive use of smart phones, there is a need for developing applications that automatically recognizes the genre, era, artist, and identity of paintings for tourism and museum industry.

1.3 Related Previous work

There has been very little work done in the area of automated fine-art classification. Most of the work done in the problem of paintings classification utilizes low-level features such as color, shades, texture and edges. Lombardi et al. [7] presented a comprehensive study of the performance of such features for paintings classification. Several formal elements of paintings such as color, line, textures etc. are being surveyed for their accuracy in classification of paintings by painter. Several supervised and unsupervised techniques are being used for classification, visualization, and evaluation including k-Nearest Neighbor, Hierarchical Clustering, Self-Organizing Maps, and Multidimensional Scaling.

There are many more studies which have reported the results of low-level features applied to painting classification task. Jana Zujovic et al. [8] conducts a similar study of color and grey level features used for classification of painting genre. Various classifiers including SVM, KNN and decision trees are tested.

Sablatnig et al. [9] uses hierarchically structured classification scheme that separates classification into 3 different levels of information color, shape of region and brush strokes.

Khan et al. [10] evaluates low-level features comprising of color, texture, shape and motion to describe visual information present inside a painting. This study also applies fusion of color and shape features as local level features using Bag of Words model for the task of painter identification.

Icoglu et al. [11] implements an indexing system for paintings that represents paintings by a six-dimensional feature vector. This feature vector uses low-level features like percentage of dark pixels, luminance histogram, gradient map etc. to capture visual contents of a painting. This feature vector is tested on Non-linear SVM for its accuracy in classifying painters.

Bressan et al. [12] visualizes the relationship between painters by constructing a graph where nodes are represented by paintings and edge weights represent the degree of similarity between any two paintings.

Thus it can be observed that most of the work done so far in case of paintings focuses mainly on low-level features. Also it is worth mentioning that most of these studies work on problem of inferring the painter from the paintings. Our study will consider more intermediate and semantic level features and identify the genre of the paintings which is a much more challenging task.

Chapter 2

Classification Models Overview

This chapter gives a basic overview of various models and methods being used and implemented in the thesis.

2.1 Image Classification

Image classification is a very important branch of Computer Vision which involves classifying an unseen image among various classes. The term ‘Image Classification’ can be formally stated as **- Image classification is the assignment of an image, which is a member of a given set of abstract classes into one such class**. Thus an unseen image should be classified to the right category it belongs to. With large number of pictures and videos available on internet, the problem of image classification has achieved a very important place in tasks of indexing and fast retrieval. With such a rich database of images, it is nearly impossible to accomplish this task manually. Thus these days a lot of research is being done to come up with models that can perform image classification efficiently and accurately.

Image classification combines many techniques from the fields of machine learning and image processing. There are two very important aspects of an image classifier –

1 Machine Learning model used

2 Features being used to represent the image

An image classifier typically applies already existing machine learning models to learn the patterns in an image and utilizes inference techniques to make classification decisions. Many important models like Neural Networks, SVM, decision trees, Naïve Bayes Classifier, KN clustering etc, have been studied and applied for the task of image classification. The type of machine learning model to be used greatly depends on the application and type of images to be processed.

Machine learning models are mathematical models which work on data typically represented in mathematical vector forms. Thus it is an essential step to represent an image by a vector which can be easily consumed by a machine learning model. Feature representation is a way to convert an image into a vector form. It can also be said that, a feature vector summarizes one or more properties present in an image into a mathematical form. This feature property can be utilized by machine learning models for recognition and classification tasks. The feature to be used is greatly determined by the type of images that are used and properties which can easily differentiate one image from another. For example, if image classes greatly differ in color than using color features can give us better results. Similarly, if image classes differ significantly by the edges present, features summarizing edge information can greatly improve our results.

Image classification involves some very important steps which include –

1) Image is pre-processed so that all images achieve uniformity and descriptors can be easily applied to the training images. Pre-processing mainly involves re-sizing of images, adjustment of pixels locations and values and edge and possibly contrast enhancement.
2) Representation of training images using a descriptor. The kind of descriptor used greatly determines the accuracy and performance of the classifier. There are different descriptors which can be local or global level. The type of application determines the descriptor to be used.
3) Training of a classifier on the training data represented using descriptors. Training can be supervised or unsupervised. In supervised training, the class labels of training data is known beforehand. On the other hand, in unsupervised learning class labels are unknown

and classifier forms clusters among data which can be used to classify a test data to one of the cluster.

4) Given a test image and trained classifier a test image is pre-processed and represented by the same descriptor being used during training. Classifier is run on the test image to infer it’s class.

2.2 Previous Work Done

There has been a lot of work done where machine learning models have been applied as classifiers for images. (Lu et al. [1],Lu et al. [2], Blei et al. [3], Liu et al. [4], Grauman et al. [5]) highlights some of the very important work done in this particular area. Various machine learning models like SVM, Neural Networks, k-Nearest Neighbor, Decision trees etc. have been very popular in their applications as image classifiers. Machine learning models are applied on a representation of an image, also known as a descriptor. Descriptor of an image is a mathematical representation of certain attributes such as color, content, texture etc. present in an image. These attributes can be learned by a classifier for further classifying new unknown images. Lot of studies in the area of image classification are based on applying different machine learning models as classifiers on the images represented using various descriptors.

Descriptors being employed in image classification can be classified into two main categories- local and global level features. Shyu et al. [6] conducts the performance evaluation of these two categories of descriptors. Some very popular descriptors extract the attributes like edges, color, texture, spatial layout and content. Research is being done to make descriptors compact and efficient.

2.3 Models Used

As discussed earlier we will perform comparative evaluation of 3 different models –

- Discriminative Semantic Model
- Discriminative Bag of Words (BoW) Model
- Generative BoW Model

These three models differ by two aspects- the classifier employed and the level of features extracted. **Table 1**. Shows the details of all three models –

Table 1. Model details

Model	**Classifier**	**Level of Features**
Discriminative Semantic Model	Discriminative Classifier (Ex. SVM)	Semantic-Level features (Ex. Classeme)
Discriminative BoW Model	Discriminative Classifier (Ex. SVM)	Intermediate-Level features using BoW technique (Ex. CSIFT, Opponent SIFT)
Generative BoW Model	Generative Classifier (Ex. LDA)	Intermediate-Level features using BoW technique (Ex. CSIFT, Opponent SIFT)

As can be observed from the table, our study is focused on Intermediate and Semantic level features to represent the painting. **Discriminative Semantic Model** uses Support Vector Machines (SVM) (Chang et al. [16]) classifier employed on Classemes (Toressani et al. [17]) which are semantic-level features. Classeme is a descriptor which is the output of a number of weakly trained object category classifiers. The output of a trained object category classifier specifies the probability of existence of that object class in the image. Thus, Classeme descriptor summarizes the object categories present in an image. Two images are distinguished based on the object categories they are composed of.

The **Discriminative BoW model** makes use of BoW technique (Csurka et al. [18]) and local level descriptors like Color Scale-invariant feature transform (SIFT) (Hakim et al. [19]) and Opponent SIFT (Sande et al. [20]) for representing images by a feature vector. Local level descriptors are used to represent local regions in an image by their respective descriptors. Thus whole image can be described as a collection of such local regional vectors instead of one global vector summarizing the whole image. To get an intermediate feature vector for the image, BoW technique is applied. The BoW model attempts to form fixed number of clusters from the collection of local regional vectors of all the images using a clustering algorithm such as K-Means clustering. This set of clusters constitutes the codebook for the image data set. All the local regions present in all of the images are substituted by their respective nearest codebook vector. Thus, every image is described by sequence of codebook vectors and can further be assigned a fixed-length histogram vector which defines how many times a particular codebook vector appeared in an image. Such a histogram vector can be used as an intermediate feature for an image.

Generative BoW model applies a Generative machine learning model on an image represented using BoW paradigm and local-level features. In our study, Latent Dirichlet Allocation Topic model (Blei et al. [21]) is employed as a generative model for the classification task. Topic models have characteristic property of capturing the hidden thematic structure of a document. Topic models have two very important parameters – topic distribution for each document and word distribution per topic. LDA is a very popular topic model that uses dirichlet parameter to determine the topic distribution for each document. Usage of LDA model involves two steps – 1) Parameter Estimation from the training data, 2) Inferring topic distribution for the new unknown data. In the experiments performed on LDA, a separate LDA model is trained for each category of painting. Given a new test painting, each category LDA model is used to infer the likelihood of painting belonging to that particular category. Category achieving highest likelihood is assigned as the predicted category.

In the next few chapters we will discuss the components of each of these models in much more details.

Chapter 3

Discriminative Semantic Model

The Discriminative semantic model applies a discriminative machine learning model on the features capturing semantic level information present in a painting. In the thesis Support Vector Machines (SVM) is used as a discriminative model and Classeme is being explored to summarize the semantic information present in a painting. Following are the steps which need to be followed to apply Discriminative Semantic Model for the task of painting classification –

- Given a set of Training paintings for each category, represent each item using Classeme descriptor. This will give us a vector of dimension 2569 for each painting.
- Train SVM classifiers on the training Classeme feature vectors. SVM will be trained for each class in a one vs. one manner.
- Given an unknown painting represent it using the Classeme vector.
- A Classeme vector is feed into SVM to classify it among one of the classes.

3.1 Support Vector Machines

Support Vector Machines is a set of related supervised learning methods that analyze data and recognize patterns and are widely used for classification and regression analysis. SVM basically constructs a hyperplane in a high dimensional feature space which is used for classification. A good separation is achieved by a hyperplane that has the largest distance to nearest training data point of any class. For non-linearly separable data set points, SVM maps finite space to a much higher dimensional space thus making separation easier in new space.

The mappings are defined in terms of Kernel function K(x,y). Kernel function ensures that the mappings are not computationally intensive and dot products may be computed easily.

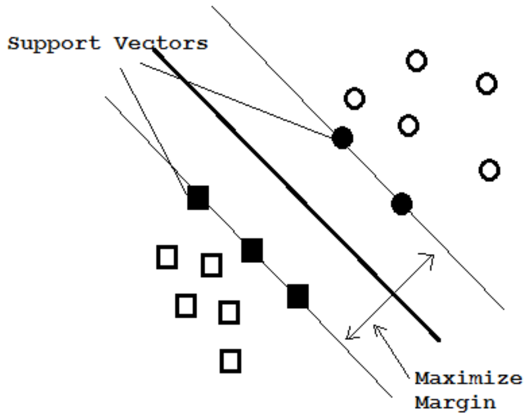

Figure 3. Support Vector Machines

The hyperplane in higher dimensional space constitute the set of points whose inner product with the vectors in that space is constant. SVM tries to find out the optimal hyperplane which depends on Support Vectors present in the space. Support Vectors are the data points that lie closest to the decision surface and are most difficult to classify. The SVM algorithm calculates the hyperplane such that the margin around the plane is maximized. **Figure 3** displays an example. Such a hyperplane is known as **maximum-margin hyperplane**. This optimal decision boundary can be utilized to make classification decisions. Cortes et al. [25] provides more details of the process.

3.2 Semantic Level features

Semantic level features became very popular with the advent of content-based image retrieval systems (CBIR). Many CBIR systems are being developed these days. To improve the retrieval accuracy of CBIR systems, research focus has been shifted from designing sophisticated low-level features to reducing the ‘semantic gap’ between the visual features and the richness of human semantics. Almost all of these systems are built on the premise that an image can be represented using global summarized vectors known as semantic level features. Such CBIR systems measure the similarity between images in a way human being would perceive or recognize. Global semantic features have the ability to generalize an entire image with a vector representing the contents of the image. These semantic content representation vectors are used to evaluate the similarity between two images.

Semantic features can be derived from the low level features. As explained in Eidenberger et al. [22] semantic features can be defined based on the lower-levels descriptors by using additional modeling, statistical and domain knowledge. Vogel et. al. in [23] presents an image representation that involves the semantic modeling. The basic idea of semantic modeling is to classify the local image regions into various semantic concepts. In case of natural scenes it can be rock, water or foliage. This study also discusses important requirements for successful semantic image representation. Liu et. al. in [24] provides a comprehensive study of five major techniques utilizing semantic level features and their applications in CBIR systems.

3.3 Classeme Descriptors

The Classemes are semantic-level descriptors that capture the visual contents present in an image. This summarized vector can be used for retrieval and classification tasks. As described in Torresani et al. [17], Classeme vector comprises the output of various weekly trained object category classifiers. Given a new image, a trained object category classifier can provide the probability value for that category object being present in the image. Thus, a feature vector comprising such probability values for each pre-defined category provides valuable information about the semantic classes present in the image. For example, a scenic view can comprise of visual classes like sky, mountains, trees etc. Classeme feature vector for such a scene will comprise higher probability values for the object classes sky, mountains and trees. These classifier values can be used to differentiate two images for the task of classification. Classeme generation has two main steps –

1. **Learning of pre-defined classifiers** –This step begins with drawing of training images out of a set of N categories. For each category 'i', the training samples are used to train a classifier Φ_i. For any new image 'I', classifier Φ_i generates a value denoted by $\Phi_i(I)$. Thus an image' I' can be represented by a Classeme vector $[\Phi_1(I),\Phi_2(I),\ldots..\Phi_N(I)]$.
2. **Application of Classeme on a new image** – Given a new image, it can be represented by a Classeme vector. This classeme feature vector can be feed into a classifier as a training input.

After the first step, training images used for basic classifiers can be discarded. The classes selected for Classeme training includes primarily of visual concepts such as trees, sky, person etc. plus abstract concepts like person walking. Classeme selects concepts from Large Scale Concept Ontology for Multimedia (LSCOM) [55]. Out of 2832 unique categories, 97 are removed as they denote abstract group of other categories. Thus, 2659 categories are used for Classeme generation. One more step which is not mentioned above involves dimensionality reduction of the feature vector. This is done by allotting a cross-validation score to each component of the vector of size 'N' and arranging them in an increasing order. To reduce the dimensionality to 'd' last 'N-d' components can be taken out of consideration.

Chapter 4

Discriminative Bag of Words Model

The Discriminative Bag of Words (BoW) model uses Support Vector Machines (SVM) discriminative classifier and trains it on the intermediate-level representation of paintings generated using BoW methodology. Discriminative BoW model involves following steps –

1) Generate interest points for all the training images.
2) Represent the interest points using local-level features like Color Scale-invariant Feature Transform (CSIFT) and Opponent-SIFT (OSIFT).
3) Form codebook using a clustering algorithm such as K-means clustering.
4) Generate intermediate-level representations for each image using codebook. It can be a histogram of regions present in an image.
5) Train discriminative classifier SVM on the intermediate training feature vectors.
6) For classification, trained SVM is applied on the BoW feature vector of an unknown image.

Rest of the topics will describe above steps and components in much more details.

4.1 Local level features - CSIFT & Opponent-SIFT

The local features denote descriptors of localized image neighborhood centered around many points of interests. Instead of one global feature vector summarizing the whole image, descriptors such as SIFT and CSIFT are applied on various local regions of the image. Thus, generation of local level features involves two steps – 1) Selection of local regions, 2) Application of SIFT and CSIFT descriptors on these regions. Once the local features at various regions are calculated, these can be matched across images. Such a distance metric can be used to devise a heuristic procedure for determining when a pair of features is considered a match, e. g. by using a distance threshold. The matching procedure may also utilize other constraints such as the geometric relationships among the interest points if the object is known to be rigid. The use of local-features also omits the need of segmentation of the objects from the image background.

The CSIFT and Opponent SIFT descriptors are based on the scale invariant SIFT descriptors. The CSIFT (Hakim et al. [19]) is based on the color invariant property applied to SIFT described in (Lowe et al. [26]). The SIFT is invariant to image scale, rotation, affine distortion and illumination. It uses edge orientations to define a local region and also utilizes the gradient of an image, derivative of which cancels out offsets and thus making it shift-invariant. Also, the SIFT descriptor is normalized and hence is also immune to gradient magnitude changes. CSIFT extends SIFT making it color invariant.

As described in Hakim et al. [20], the Opponent-SIFT descriptor makes use of opponent histogram. The opponent histogram is a combination of three channels of Opponent space - $O_{1,}$ O_2 and O_3. O_3 consists of the intensity information while the color information is held by channels O_1 and O_2. In opponent SIFT, a descriptor is computed for each channel and the combined feature vector is used to represent an image.

CSIFT differs from Opponent sift in a way that it removes all the intensity information available in channels O_1 and O_2 and thus adding invariance to the intensity changes. The CSIFT descriptor uses color invariant normalized color space O_1/O_3 and O_2/O_3 as input to SIFT. Due to the denominator factor, CSIFT is invariant to any light intensity changes.

4.2 Bag of Words model for Image Classification

The BoW (Csurka et al. [18]) is a very popular model used to represent documents where order of words is totally ignored. In the BoW model, each document is considered as a bag of words where only content matters, irrespective of the order of words. Using this model, images can be very well related to the documents. An image can be seen as a collection of similar regions. Such regions represented using local descriptors can be considered as words for an image. Thus when we represent the images using BoW model, the properties like spatial layout and relative arrangements of objects in an image are not captured. The BoW model provides a very useful framework for applying local-level descriptors to the task of image classification. The images in a BoW model can be represented by a co-occurrence matrix M where each element m_{ij} represents count of word i in a document j.

Several steps are required to be performed to get a BoW representation of an image. **Figure 4**. depicts various steps involved in BoW representation of an image.

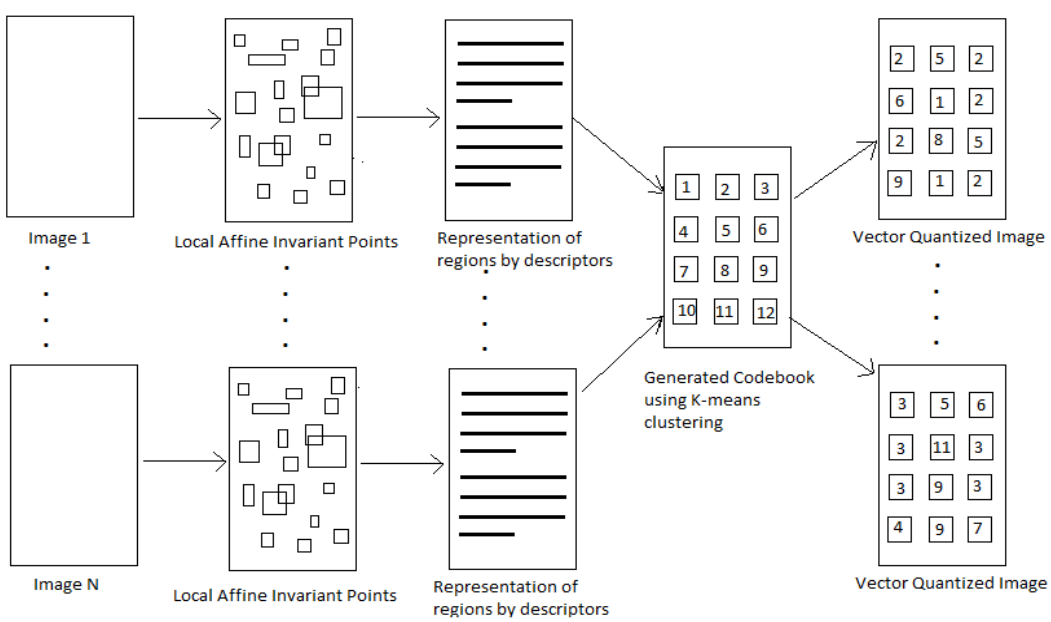

Figure 4. Bow representation of an image

4.2.1 Extraction of interest points

The interest point detector detects local regions or patches such as edges, corners and square regions in an image. These selected points are equivalent to the words in an image. Some of the very important detectors includes –

- **Harris affine region detector –** This detector uses the Harris-Laplace detector to locate the points of interests. These interest points share the property that they will be compatible across images taken at different viewpoints. More details can be referred from Mikolajcyk et al. [29].
- **Hessian affine region detector –** This detector is quite similar to the Harris detector. Like Harris detector, it also locates affine-invariant regions in an image. Hessian detector chooses interest points based on the Hessian Matrix. Mikolajcyk et al. [29] provides more details of the process.
- **Kadir-Brady saliency detector –** The Kadir-Brady detector finds out the unique points in an image which are also invariant to transformations and illumination. Details of this are explained in Kadir et al. [30].

Mikolajczyk et. al. [28] provides a comprehensive comparison of various commonly used detectors.

4.2.2 Representation of interest points using descriptors

This step deals with the numerical representation of the regions located by the previous step. The local level descriptors like CSFIT and Opponent-SIFT discussed earlier can be applied to interest regions to get a feature vector representation of these. Thus, our image now can be represented using collection of these vectors. This can easily be related to a document represented by a collection of words.

4.2.3 Codebook Formation

Next very important step is to build a codebook from the local region descriptors generated by the previous step. Descriptors of all the points present in all of the images are set as input to a clustering algorithm such as K-means clustering. Also the size of Codebook has to be decided before execution of clustering algorithm. The clustering algorithm provides us a collection of words where each word is represented by the descriptor used earlier. This collection of descriptors also represents the total words or vocabulary of our data-set.

4.2.4 Vector quantization of Images

Once we have the codebook ready, images can be represented with a quantized vector of these words. This can be performed by denoting each region descriptor present in the image with a nearest codeword in the codebook. Clustering algorithms can be utilized for this task. This way all the interest regions present in an image are assigned a codeword present in the codebook.

4.2.5 Representation of image by histogram of codewords

The quantized vector we received from the last step is used to generate a histogram of words in codebook. Thus each element of this histogram represents the number of times that particular codeword has appeared in the image. Now each image can be represented by this fixed length histogram vector. The length of the vector is equal to the number of words present in the codebook. A classifier can be easily trained on such image histogram vectors.

4.3 Previous work based on BoW technique

In recent years, the BoW models have appeared to be very popular for the tasks of image classification. Various studies have utilized this model to represent local level features for the images.

- Jiang et al. [50] evaluates BoW model for the task of scene classification. Various features such as Mutual information, document frequency etc. are applied on various regions of the training images.
- Sivic et al. [51] explored BoW model for the task of object and scene retrieval by searching all the occurrences of an object in the video stream. SIFT features are used with PLSA model.
- Csurka et al. [18] applies BoW model with SVM classifier.
- Larlus et al. [52] applies BoW model for object categorization in an image using BoW and Gaussian-Multinomial LDA.
- Bosch et al. [53] applies local shape features on various regions of the image using BoW model and thus performs the task of object categorization.

Chapter 5

Generative Bag of Words Model

The Generative BoW model is an automatic classifier model that employs a generative topic model known as Latent Dirichlet Allocation (LDA) and Bag of Words (BoW) representation as an intermediate level description for the paintings. For classification purpose, a technique similar to the one applied in Fei-Fei et al. [13] is implemented for the paintings data set. Generative BoW model application involves following steps –

1. Generate BoW representation for all the training paintings.
2. For each class, use its corresponding training samples and estimate each class's LDA parameters. Thus, there will be a trained LDA model for each category of paintings.
3. Given a new test painting, represent it using already generated codebook.
4. For each category, use LDA model to infer the likelihood of the test painting. Category achieving highest likelihood score is the predicted class of the painting.

5.1 Topic Models

With millions of documents making their place into digital library, the task of automatic management and organization of these documents holds a lot of importance. For automatic and fast retrieval of documents we need some kind of index into our document library**.** Topic model (Steyvers et al. [31]) is a class of probabilistic models which provides a way to extract the underlying and hidden semantic structure present in a document. Topic models can be applied to documents, newspaper articles, scientific journals, images, emails and lot of other kinds of media.

Topic models are based on the idea that a document can be represented by a hidden topic distribution; where each topic in turn is assumed to be a mixture of words. Based on these two assumptions, a topic model can be very well described as a generative model for the construction of a new document. The Generative process for modeling a document can be described as - first select a topic distribution for the document, than for each word in the document its topic is selected using the topic distribution, and the word distribution for that particular topic is utilized to construct the word. Thus any two documents are differentiated by the topic distribution present.

The LDA as described in Blei et al. [21] is a very important and widely used topic model. Next few sections describe details about LDA, its geometrical graph model, generative process and parameter estimation technique.

5.2 Topic Model LDA

LDA is a topic model goal of which is to automatically identify the topic structure hidden in a collection of documents. LDA model works in an unsupervised manner. Given a collection of documents, LDA model estimates the hidden topic distribution in every document and the word distribution for each topic.

Figure 5 shows the graphical model of LDA. It is assumed that there are D documents and N words per document. As showed in the graphical model, parameter 'ϴ' defines the topic distribution for each document, 'β' denotes word distribution for each topic, 'α' is the dirichlet parameter that determines topic distribution parameter ϴ and 'z' is the topic for each word 'w' in a document.

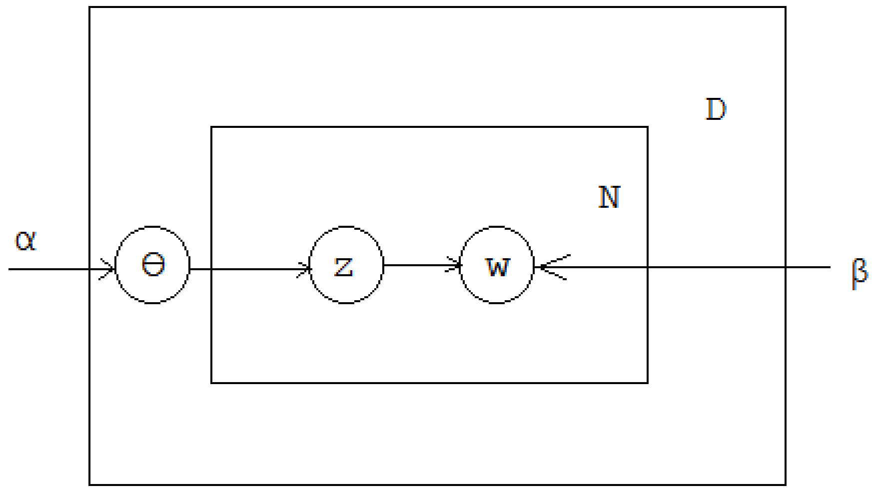

Figure 5. LDA graph model

In LDA, β, Θ, α and Z are considered to be the hidden values. Only known parameter is W which is the contents of the document. The hidden variables determine the topical structure exhibited by the documents. The posterior estimates of these hidden variables given observed data can be used to identify the topics present in the documents and thus provides indexing into the collection.

5.2.1 Generative process for LDA

The LDA model can be very well viewed as a generative process using which new documents can be constructed. To describe the generative process we set number of topics to be 'K'. A particular document is constructed word by word. For a document 'd' –

1. Choose topic distribution 'Θ_d' which is determined by the dirichlet function of parameter 'α'

$$\Theta_d \sim Dir(\alpha)$$

2. For each word 'w_n' in document 'd' choose a topic 'k' from the Multinomial distribution of 'Θ_d'.

$$k \sim Mul(\Theta_d)$$

3. For a topic 'k', use distribution 'β_k' to draw out a word 'w_n'.

$$w_n \sim Mul(\beta_k)$$

4. Repeat steps 2 and 3 for each word.

Thus a new document can be synthesized word by word using the thematic structure parameters β, Θ and α

5.2.2 Application of LDA on images

Figure 6 presents the whole procedure for applying LDA for the task of image classification

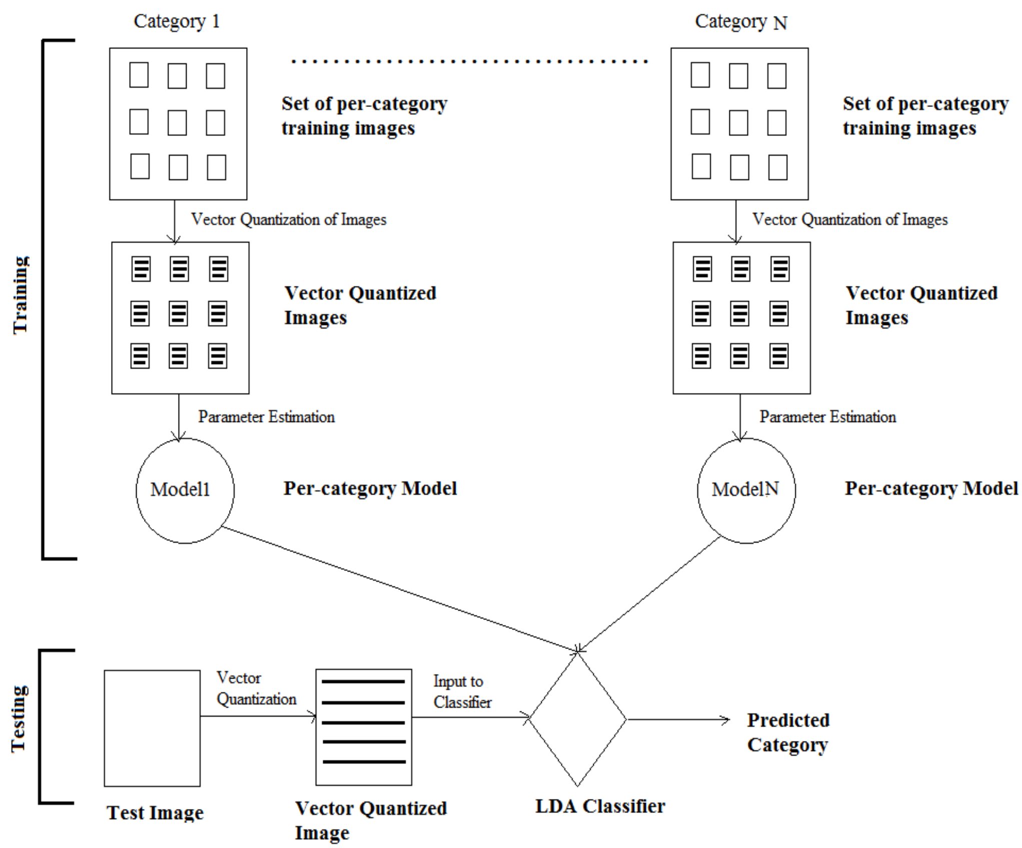

Figure 6. Application of LDA on images

Topic models usage in the case of images is very similar to that for documents. Documents can very easily be visualized as a collection of words; images on the other hand can be augmented with such a structure using Bag of Words model. As explained in last chapter, bag of words model can describe an image as a collection of local regional vectors. Such local regional vectors can be referred as words for LDA model. We have used LDA as a generative model for the task of paintings classification. Initially LDA model is estimated for each category. Then given a new test image, we use inference techniques to find out the likelihood of a painting belonging to a particular class. Following are the two phases of the process –

Training Phase – It includes –

1. Sort out the training images by categories.
2. Represent each training image by a quantized vector using Bag of Words model. It involves all the steps mentioned in **section 4.2** including representation of an image by descriptors, formation of codebook and generating word histogram for each image using K-means clustering. In the end, all the training images are represented by the intermediate-level BoW feature vectors.
3. The parameters are estimated for each category LDA model. As described in Blei et al. [21], finding the posterior distribution of the hidden variables given a document is intractable due to the coupling between β and Ө. So an approximation technique known as Variational Inference can be applied for the hidden parameters estimation. Variational Inference makes use of Jensen's inequality to obtain lower bound on likelihood. In the variational inference technique, simple modifications that involve removal of some of the edges or nodes are applied to the original graphical model. Thus the LDA graphical model depicted in **Figure 5** can be simplified by eliminating edges between Ө, z and w. Such a simplification process removes coupling between Ө and β and thus makes estimation process tractable. **Figure 7** shows modified graphical model.

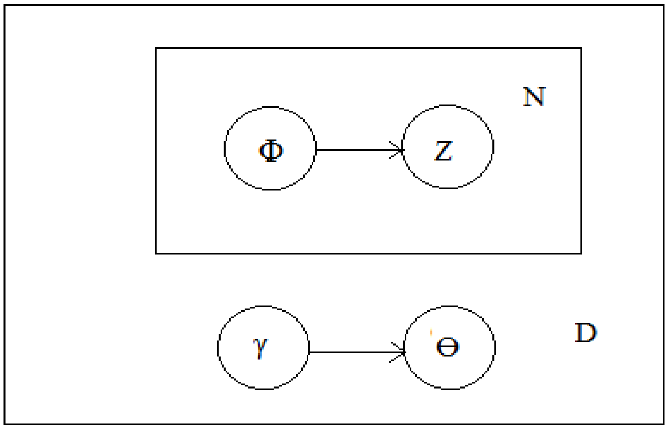

Figure 7. Modified LDA graphical model

As can be seen, the modified graph has lost the coupling between Θ and β. The new variational parameters γ and Φ have been introduced into the model. Here γ is a Dirichlet parameter and ($\Phi_1,\ldots, \Phi_N$) is a set of multinomial parameters for each word. Next step is to approximate the values of γ and Φ by minimizing the Kullback-Leibler (KL) divergence between variational distribution and original distribution. The full details of the methods can be referred from Blei et al. [21].

4. At the end of Training phase we will have estimated LDA models for each category which can be utilized during testing of an unknown painting.

Testing Phase – Given an unknown painting, testing phase involves –

1. Using each category model and its estimated parameter β, infer the value of Θ for the new unknown painting.
2. From the estimated parameters, likelihood values for each category can be inferred. Class achieving the highest likelihood is the predicted class for the sample test painting.

5.3 Previous Work Based on Topic Models

Use of Topic models is very common in the applications that involve classification, indexing and the retrieval of a large set of unclassified documents. Recently there has been a lot of research work done for applying topic models in the field of image classification and recognition.

- Fei-Fei et al. [13] applies LDA to the task of scene category recognition in the case of ordinary images. BoW model is applied to Local level features SIFT to get an intermediate representation of the images. Further, LDA model for each given category is learned which is then later used to infer the category of an unknown scene present in an image.

- Sivic et al. [14] discovers the object categories in a set of unlabelled images using Probabilistic Latent Semantic Analysis (pLSA) (Hofmann et al. [15]) topic model. The BoW intermediate representation is utilized by pLSA to predict the topic distribution present in the data-set which can be utilized to perform segmentation in an unsupervised manner.

- Burns et al. [40] applies topic models for the task of unsupervised segmentation of large set of scanned documents. LDA is being utilized for this purpose.

- Yanai et al. [41] implements PLSA model for the task of keyword association with a large set of web based images.
- Bart et al. [44] applies hierarchical topic model to create a tool TAX which learns and generates Taxonomy of the various images in an unsupervised manner.
- Li et al. [45] performs a similar work of hierarchical clustering for a large set of images. Images are organized in a general-to specific relationship.

Chapter 6

Experimental setup

Experiments performed in the thesis are directed towards comparative study of three models used for the automatic task of painting style classification. This chapter provides details of the data set used and the experimental settings applied at the time of performing experiments for each of the model. All three models are measured using classification accuracy as a parameter. Confusion matrices are calculated for each technique to get better insights into the classification results.

6.1 Data Set

We performed classification experiments for seven fine-art genres namely – *Renaissance, Baroque, Popart, Expressionism, Impressionism, Cubism and Abstract*. Total of 70 paintings per painting style are selected for the purpose of experiments. The paintings are taken from [32] which is a large repository of fine art paintings. 5-fold cross-validation is performed to randomly select 20% (14 paintings) as test data for each class. Results are further compared for each fold.

6.2 Codebook formation settings

As explained in chapter 4, Codebook formation is a very important step in the implementation of Generative and Discriminative Bag of Words (BoW) models. Codebook settings can greatly influence the vector representation of the paintings. For codebook generation, Harrisp Laplace detector is run for the purpose of isolating interest points. The Harris Laplace salient point detector uses a Harris corner detector and Laplacian for scale selection. These interest points are invariant to rotation, scale, illumination and noise.

The code available at Vl Feat. [33] is being utilized for the tasks involving codebook generation. To improve the performance, number of interest points per painting are restricted to 3000. The K-means algorithm is configured to build a codebook of size 600 clusters. These clusters constitute the vocabulary of the codebook.

6.3 Support Vector Machines settings

The Support Vector Machines (SVM) classifier is trained for both Discriminative semantic and BoW model. For SVM, Radial Basis function (RBF) is being used as a Kernel. The RBF kernel has the property that it nonlinearly maps samples into a high dimensional space, thus it can handle the case where relation between class labels and features is non-linear. Also the number of hyperparameters that influence the complexity of model selection is lower than that of a polynomial Kernel.

There are two parameters for a RBF Kernel - C and γ. To determine these parameters, grid search algorithm implemented by Chang et al. [16] is being used. The grid search algorithm uses the cross-validation technique to pick up the optimum parameter values. The K-fold Cross-validation involves division of training set into K subsets. Every subset is tested once using the classifier trained on rest of the K-1 subsets. Also, this process is preceded by the scaling of data set descriptors. Scaling avoids attributes in greater numeric ranges dominating those in smaller ones.

6.4 Latent Dirichelet Allocation Settings

For experiments with Latent Dirichlet Allocation (LDA) David Beli’s C-code [34] is being used for the task of parameter estimation and inference. The C-code uses the variational inference technique which tries to estimate the parameters ‘β’ and ‘Θ’. The code also provides functionality for inference which can be used for classifying a new unknown painting.

For parameter estimation, alpha is set to 0.1 and LDA code is set to estimate the value of α during estimation process. Also the number of topics ‘T’ is equal to the number of categories which is 7.

Chapter 7

Results

As described in previous chapters, all three models are tested for their accuracy of classifying paintings. To get more detailed measure of the results, confusion matrices displaying results for each category are generated. This chapter will summarize all the results of the experiments being performed. As explained in last chapter, 5-fold cross validation is performed and the average of confusion matrices for all the folds is been taken as final result.

7.1 Discriminative Semantic model

Table 2 displays overall results of the Discriminative Semantic model which applies Support Vector Machines (SVM) on the semantic-level features 'Classemes'.

Table 2 Discriminative Semantic Model

Overall Accuracy	65.4%		Number of correctly classified paintings		64/98	
Confusion Matrix							
ACCURACY(%)	Baroque	Abstract	Renaissance	Popart	Expressionism	Impressionism	Cubism
Baroque	87.5	0	14.3	0	5.3	17.8	1.78
Abstract	0	64	0	7.1	7.1	1.8	1.9
Renaissance	5.4	0	64.3	5.35	14.3	3.5	0
Popart	0	1.78	1.8	73.1	0	3.5	1.8
Expressionism	1.8	20.2	7.1	3.6	48.2	17.8	12.9
Impressionism	5.36	8.	9	5.3	17.8	48.2	9.2
Cubism	0	6	3.5	5.3	7.1	7.1	72.4

As can be seen, the overall accuracy for the model comes out to be 65.4%. **Table 2** also displays a confusion matrix for all the seven classes of paintings. From the results it can be derived that model behaves well for Baroque, Popart and Cubism categories. While it proved incapable of distinguishing between Expressionism and Impressionism styles.

7.2 Discriminative BoW Model

These set of experiments involved application of SVM classifier on two local level features: Color SIFT and Opponent SIFT. **Table 3** and **4** displays accuracy results for these experiments.

Table 3 Classification results for Discriminative BoW using CSIFT

Overall Accuracy	48.47%		Number of correctly classified paintings		47/98		
Confusion Matrix							
ACCURACY(%)	Baroque	Abstract	Renaissance	Popart	Expressionism	Impressionism	Cubism
Baroque	71.4	0	12.9	0	8.5	17.1	0
Abstract	0	48	5.8	10	8.5	5.7	7.1
Renaissance	18.6	6.7	41.4	0	5.8	9.3	18.5
Popart	0	15	0	70	11.5	9.3	15.7
Expressionism	0	15	18.6	2.8	28.5	12.9	13
Impressionism	8.5	8.6	3.7	8.6	17.2	45.7	11.4
Cubism	1.5	6.7	17.6	8.6	20	0	34.3

Table 4 Classification results for Discriminative BoW using OSIFT

Overall Accuracy	56.7%		Number of correctly classified paintings		55/98		
Confusion Matrix							
ACCURACY(%)	Baroque	Abstract	Renaissance	Popart	Expressionism	Impressionism	Cubism
Baroque	82.1	0	10.7	0	14.3	17.9	3.6
Abstract	0	54.2	3.6	7.1	7.1	3.6	7.1
Renaissance	3.6	0	64.3	3.6	21	0	7.1
Popart	0	12.5	3.6	75	0	0	17.9
Expressionism	0	16.7	0	3.6	36	10.7	28.6
Impressionism	14.3	8.33	7.2	3.6	10.7	57.1	7.1
Cubism	0	4.2	10.8	7.1	14.3	10.7	28.6

As can be inferred from the tables, Discriminative BoW using CSIFT achieves overall accuracy of 48.5% and with OSIFT it is 56.7%. Thus classification of styles using OSIFT performs better than CSIFT features. Also, Baroque style once again got maximum correct predictions. While confusion between Expressionism and Impressionism styles still exists.

7.3 Generative BoW Model

This methodology constitutes generative model that applies LDA on CISFT and OSIFT features. **Table 5** and **6** summarizes the results for Generative BoW on CSIFT and OSIFT. Overall accuracy achieved with CSIFT is 49% while with OSIFT it is 50.3 %. Thus unlike Discriminative models, Generative models achieve almost equivalent performance with CSIFT and OSIFT features.

Table 5 Classification results for Generative BoW using CSIFT

Overall Accuracy	49%		Number of correctly classified paintings		48/98		
Confusion Matrix							
ACCURACY(%)	Baroque	Abstract	Renaissance	Popart	Expressionism	Impressionism	Cubism
Baroque	86.6	0	14.3	0	14.3	7.1	7.1
Abstract	0	58.3	7.1	26.6	0	7.1	14.3
Renaissance	6.6	8.3	42.8	20	14.3	0	7.1
Popart	0	0	7.1	13.3	0	0	14.3
Expressionism	0	8.3	7.1	6.6	36	14.3	7.1
Impressionism	6.6	25	14.3	13.3	21.4	71.4	14.3
Cubism	0	0	7.1	20	14.3	0	35.7

Table 6 Classification results for Generative BoW using OSIFT

Overall Accuracy	50.3%		Number of correctly classified paintings		49/98		
Confusion Matrix							
ACCURACY(%)	Baroque	Abstract	Renaissance	Popart	Expressionism	Impressionism	Cubism
Baroque	75.5	0	14.3	0	3.6	10.7	7.1
Abstract	0	62.5	3.5	27.3	3.6	3.6	0
Renaissance	7.1	4.2	39.2	3.3	7.1	3.6	10.7
Popart	0	8.3	0	28	3.6	0	7.1
Expressionism	7.1	0	17.8	14	36	3.6	10.7
Impressionism	10.2	25	10.7	10.2	32	68	21.4
Cubism	0	0	14.3	16.9	14.3	10.7	42.9

7.4 Overall Result Summary

Table 7 and **Figure 8** displays summarized results of the comparative study performed. As can be examined the Discriminative model with Semantic-level features achieved the highest accuracy followed by Discriminative BoW with OSIFT, Generative BoW with OSIFT, Generative BoW with CSIFT and Discriminative BoW CSIFT. **Table 7** also shows variance involved in the accuracy percentages.

Also it can be deduced from the results that both Discriminative and Generative BoW models achieved comparable accuracy, while Discriminative Semantic model outperforms both BoW models.

Table 7 Overall Summary

Overall Summarized Results				
Discriminative Semantic Model	Discriminative BoW CSIFT	Discriminative BoW OSIFT	Generative BoW CSIFT	Generative BoW OSIFT
65.4%	48.47%	56.7%	49%	50.3%
Std 4.8%	Std 2.45%	Std 3.26%	Std 2.43%	Std 2.46

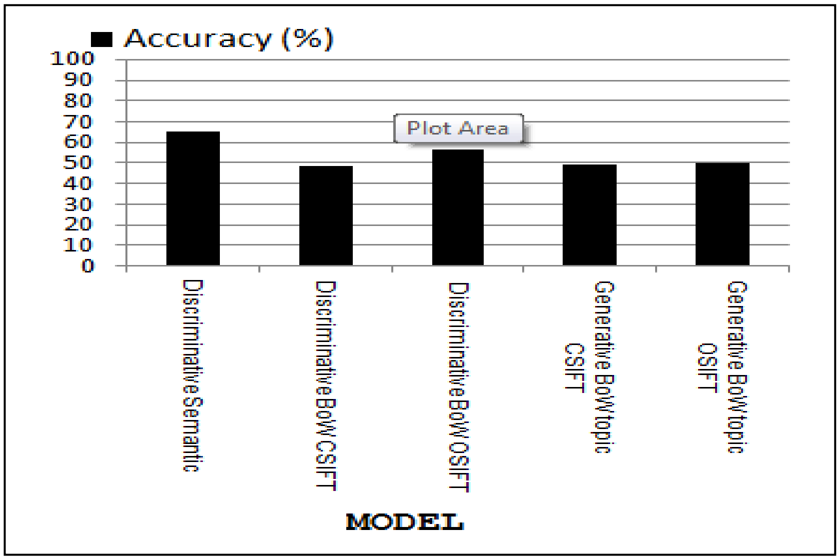

Figure 8. Performance graph for Models

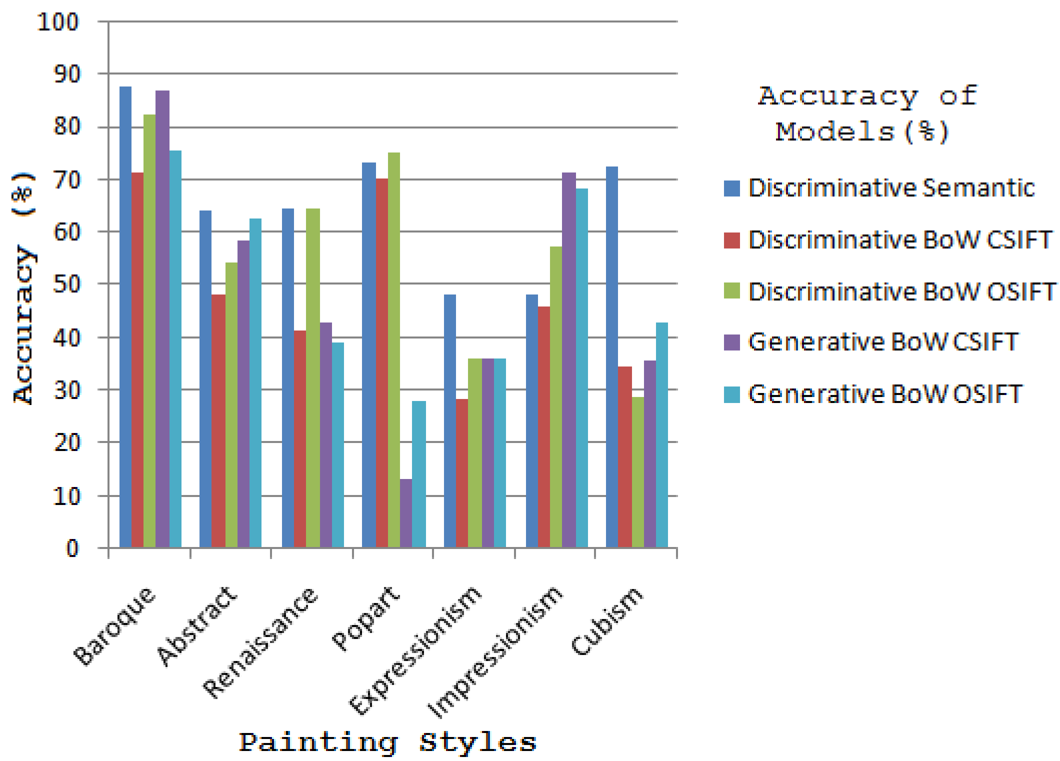

Figure 9. Per style per model Accuracy % Chart

Chart depicted in **Figure 9** highlights accuracy achieved by each model for each painting style.

By examining the results we can notice that the Baroque style is always classified with the highest accuracy in all techniques. It is also interesting to notice that the Popart genre is classified with accuracy over 70% in all the discriminative approaches while the generative approach performed poorly in that genre.

Also it is worth noting that the OSIFT features outperformed the CSIFT features in the discriminative case; however the difference is not significant in the generative case.

Chapter 8

Conclusion

In the thesis we studied the problem of paintings classification and evaluated three different models for the classification task. The study involved generative and discriminative machine learning models applied on semantic and intermediate level feature representations of the paintings. For experiments, seven categories of fine-art genres were chosen with 70 paintings per style. Chapter 7 highlights the results achieved using our study.

From the results we can derive the following conclusions-

- The semantic level features like Classemes work really well in representing paintings for the task of classification. Till now not much of the emphasis has been put on the semantic level representation of the paintings. Previous work mostly involved low level features for this task. From Table 7, it is quite evident that Discriminative Semantic model achieves highest accuracy of 65.4%. This supports the notion that semantic knowledge maintains its relevance for the representation of a painting and it is worth experimenting with various different semantic-level features for the similar tasks.
- The Generative topic models like Latent Dirichlet Allocation (LDA) can very well capture the topic distribution hidden inside a set of paintings which can be further utilized for the purpose of paintings classification. As can be referred from Table 7, accuracy achieved by Generative topic models is quite comparable to that attained by Discriminative model for the case of CSIFT; 49% and 48.7% respectively. Till now topic models have only been applied to text documents and ordinary images. The results of the experiments shows that the topic models are equally powerful when applied to a large set

of paintings and thus provide necessary motivation for applying various other similar models on much larger painting sets.

Hence, the results are in line with the hypothesis we stated in chapter 1. This encourages us to try out various other models and techniques for the painting classification task. Also there are many other aspects of painting classification that are needed to be studied and applied. For example paintings can be organized & classified in a hierarchical manner with Styles as top levels and painters information at the lower levels. Also, if it is possible to model the time-eras of the paintings, lot of valuable information can be inferred from the unknown paintings. Thus this opens a lot of opportunity for the new research involving experimental study of various other different models and features for the paintings classification task.

References

[1] Image categorization with spatial mismatch kernels Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Computer Vision by Zhiwu Lu, H. H. S. Ip Vol. 0 (2009), pp. 397-404

[2] Image categorization by learning with context and consistency Computer Vision and Pattern Recognition, IEEE Computer Society Conference on Computer Vision by Zhiwu Lu, H. H. S. Ip, Vol. 0 (2009), pp. 2719-2726

[3] Simultaneous image classification and annotation Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Vol. 0 (2009), pp. 1903-1910, doi:10.1109/CVPRW.2009.5206800 by Chong Wang, D. Blei, Fei F. Li

[4] Scene Modeling Using Co-Clustering In IEEE 11th International Conference on Computer Vision (2007) by Jingen Liu, Mubarak Shah

[5] The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features Computer Vision, IEEE International Conference on In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 2 (21 October 2005), pp. 1458-1465,doi:10.1109/ICCV.2005.239 by Kristen Grauman, Trevor Darrell

[6] C. R. Shyu, C. E. Brodley, A. C. Kak, A.Kosaka, A. Aisen, and L. Broderick, Local versus global features for content-based image retrieval, in Proc. IEEE Workshop of Content-Based Access of Image and Video Databases, Santa Barbara, CA, June 1998, pp. 30–34.

[7] T. Lombardi. The Classification of Style in Fine-art Painting. Dissertation, Pace University, 2005.

[8] J. Zujovic, L. Gandy, S. Friedman, B. Pardo, and T. N. Pappas. Classifying paintings by artistic genre: An analysis of features & classifiers. In in Proc. IEEE Workshop on Multimedia Signal Processing, 2009.

[9] R. Sablatnig, P. Kammerer, and E. Zolda. Hierarchical Classification of Paintings Using Face- and Brush Stroke Models. ICPR 1998

[10] Fahad Shahbaz Khan, Joost van de Weijer, & Maria Vanrell. (2010). Who Painted this Painting?

[11] O. Icoglu, B. Gunsel, and S. Sariel. Classification and Indexing of Paintings Based on Art Movements, 12th European Signal Processing Conference 2004, September 6-10, 2004. Vienna, Austria, pp. 749-752.

[12] M. Bressan, C. Cifarelli, and F. Perronnin. An analysis of the relationship between painters based on their work. In ICIP, 2008.

[13] L. Fei-Fei and P. Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories.IEEE Comp. Vis. Patt. Recog. 2005

[14] Josef Sivic, Bryan C. Russell, Alexei A. Efros, Andrew Zisserman, William T. Freeman "Discovering Objects and their location in images" In ICCV 2005.

[15] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. Machine Learning, 43:177–196, 2001.

[16] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[17] Lorenzo Torresani, Martin Szummer, Andrew Fitzgibbon. Efficient Object Category Recognition Using Classemes. European Conference on Computer Vision, 2010

[18] G. Csurka, C. Dance, L.X. Fan, J. Willamowski, and C. Bray (2004). "Visual categorization with bags of keypoints". Proc. of ECCV International Workshop on Statistical Learning in Computer Vision.

[19] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006, pp. 1978–1983.

[20] Koen E. A. van de Sande, Theo Gevers and Cees G. M. Snoek, Evaluating Color Descriptors for Object and Scene Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence

[21] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. Journalof Machine Learning Research, 3:993–1022, 2003.

[22] H. Eidenberger, C. Breiteneder, Semantic feature layers in content-based image retrieval: implementation of human world features, in: Proceedings of International Conference on Control, Automation, Robotics and Computer Vision, Singapore, 2002 (published on CD).

[23] Vogel, J., & Schiele, B. Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision.

[24] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 40(1):262–282, 2007.

[25] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995

[26] Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157

[27] Two bag-of-words classifiers ICCV 2005 short courses on Recognizing and Learning Object Categories .

[28] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, A comparison of affine region detectors. In IJCV 65(1/2):43-72, 2005

[29] Mikolajcyk, K. and Schmid, C. 2002. An affine invariant interest point detector. In Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada.

[30] Kadir, T., Zisserman, A. and Brady, M. An affine invariant salient region detector. Proceedings of the 8th European Conference on Computer Vision.

[31] Mark Steyvers; Tom Griffiths (2007) "Probabilistic Topic Models" In: T. Landauer, D McNamara, S. Dennis, and W. Kintsch (eds), Handbook of Latent Semantic Analysis, Psychology Press.

[32] http://www.artchive.com/

[33] http://www.vlfeat.org/

[34] David Blei LDA C-code available at http://www.cs.princeton.edu/~blei/lda-c/

[35] D. Blei and J. Lafferty. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, 2006.

[36] D. Blei and J. Lafferty. Correlated Topic Models. Neural Information Processing Systems, 2006

[37] D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. Neural Information Processing Systems 16, 2003

[38] Feng, Yansong and Mirella Lapata. 2010b. Topic models for image annotation and text illustration. In Proceedings of the NAACL HLT. Association for Computational Linguistics, Los Angeles, California, pages 831–839.

[39] Lawrence, N. D. (2000). Variational inference in probabilistic models. Doctoral dissertation, Computer Laboratory, University of Cambridge, New Museums Site, Pembroke Street, Cambridge, CB2 3QG, U.K.

[40] T. J. Burns and J. J. Corso, “Robust Unsupervised Segmentation of Degraded Document Images with Topic Models,” In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1287-1294, 2009.

[41] Keiji Yanai: Automatic Web Image Selection with a Probabilistic Latent Topic Model, Proc. of the Seventeenth International World Wide Web Conference, Poster Paper, Beijing, China (2008/04).

[42] Duangmanee Putthividhya, Hagai Thomas Attias, Srikantan S. Nagarajan: Supervised topic model for automatic image annotation. ICASSP 2010: 1894-1897

[43] D. Blei and M. Jordan. Modeling annotated data. In Proceedings of the 26th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 127–134. ACM Press, 2003.

[44] Unsupervised organization of image collections: taxonomies and beyond Evgeniy, Bart Max Welling, and Pietro Perona

[45] L.-J. Li, C. Wang, Y. Lim, D. Blei, and L. Fei-Fei. Building and using a semantivisual image hierarchy. In CVPR, 2010.

[46] Josef Sivic, Bryan C. Russell, Andrew Zisserman, William T. Freeman, Alexei A. Efros: Unsupervised discovery of visual object class hierarchies.CVPR 2008

[47] Pedro Quelhas, Jean-Marc Odobez: Natural Scene Image Modeling Using Color and Texture Visterms. CIVR 2006: 411-421

[48] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proc. ICCV, Oct 2003.

[49] R. Fergus, P. Perona, and A. Zisserman. A visual category filter for Google images. In Proc. ECCV. Springer-Verlag, May 2004.

[50] Y. G. Jiang, C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object ategorization and semantic video retrieval," in ACM CIVR, 2007

[51] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In Proceedings of ICCV, volume 2, pages 1470–1477, Nice, France, oct 2003.

[52] D. Larlus and F. Jurie. Latent mixture vocabularies for object categorization. In Proceedings of the British Machine Vision Conference, 2006.

[53] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In Proceedings of ICCV, 2007.

[54] Ng, A.Y., Jordan, M.I., 2001. On discriminative versus generative classifiers: a comparison of logistic regression and naive Bayes. Adv. Neural Inform. Process. Syst. 14, 605–610.