Detection of Forgery in Art Paintings using Machine Learning

Article · May 2017

1 author:

Sunil B. Mane
College of Engineering, Pune

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

Detection of Forgery in Art Paintings using Machine Learning

Rinku Nemade[1], Apoorva Nitsure[2], Poorwa Hirve[3], Sunil B. Mane[4]

B.Tech. Student, Department of Information Technology, College of Engineering Pune, India[1,2,3]

Associate Professor, Department of Information Technology, College of Engineering Pune, India[4]

ABSTRACT: This project aims at the identification of art paintings by using machine learning. In today's world, computers are advancing faster than ever and more and more applications of machine learning have been explored, like computer vision. We have identified this problem statement as relevant and challenging. Currently detection of forgery in museums is done by examining the painting in detail by an art expert. Ongoing research on automated art identification is limited. Through machine learning, we aim to identify if two paintings are painted by the same person. We believe that our proposed solution for detecting forgery in art paintings hold interesting applications for curators and art historians, and for connoisseurs and art lovers. Through our project, the similarity between the different artists can be found out, with the characteristics and style of the paintings identified through machine learning.

KEYWORDS: Text detection, Inpainting, Morphological operations, Connected component labelling.

I. Introduction

What is forgery? Forgery can be said to be an imitation or copy of original bank notes, art works, documents meant to be sold in the open market for monetary as well as social benefits. Paintings, film direction,sculptures, dance choreography and so on in which fake art is associated with famous artistsall culminates into forgery. In this paper we focus on forgery in art paintings. Currently experts have to physically handle the painting to check if it is an original piece of work. There are various rigorous processes for forgery detection. Experts have to apply various chemicals to check the color pigments, make sure the canvas is right and so on. However we aim to automate this work at least to act as an initial step or first pass in determining if the painting is forged or not. Machine learning solutions have been proposed in this project to aid in the forgery detection problem with output as similarity index. Advancements in machine learning and image processing have led to increased accuracy making digital detection to be the most viable option in current times. Finding the individual style of a painter or artist that is his unique factor in turn helps in the forgery identification. In this project we are extracting features from paintings by which we are determining the style of an artist to lead up to detecting forgery.

II. Scope and Objectives

Using this system, automatically classifying a digital capture will help muse-ums to organize large digital collections, and art appreciators to gain insight into the paintings. Our scope is limited in terms of the physical aspects of forgery detection as we use scanned images and machine learning.

1. Scope

Our systems scope lies in the following areas:

1. Applicable to paintings done by famous artists
2. Identifying the genre

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

3. Identifying the artists
4. Identifying Photoshopped images
5. Applicable to medical field where MRI or X-ray can be studied

2. Objectives

   1. Identifying the artist from the painting
   2. Detection of forgery by comparing styles and getting similarity
   3. Detecting fake images
   4. Identifying artists having similar painting style

III. LITERATURE SURVEY

Our references contain all the papers that we have obtained and studied in order to do this project. The main paper that are related to the steps taken to apply the project are explained in the following:

1. Detection of forgery in paintings using supervised learning[26]

Variations in image clarity in the experimental data sets were correlated with authenticity. Supervised machine learning on features derived from Hidden-Markov-Tree-modelling of the paintings wavelet coefficients has the potential to distinguish copies from originals. Visual aspects of appearance and style of the work are compared against those of the artists other works. Parameters of an HMT, computing a blur index for each painting, a number that represents the degree of (visually unnoticeable) blurriness of each image, based on the detection scheme. Hidden Markov Tree (HMT) model that captures statistical structure of the image. For each sub band, the wavelet coefficients form a quad-tree with structural local dependencies at different levels.

2. Classifying Paintings by Artistic Genre: An Analysis of Features and Classifiers[27]

This paper uses the method of grayscale, colorspace and classification of the images in order to detect authenticity of the painting. Grayscale provides information about the texture using Steerable Pyramid Decomposition which calculates mean and variance of the absolute coefficients in each sub-band which characterizes the brush strokes of the artist. Edges can also be detected by gray-level by labelling the edge pixels according to different sensitivity thresholds to identify the painting styles characterized by smudges or subtleedges. As sensitivity threshold decreases, subtle edges will be more distinct whereas clear edges will be observed at any threshold. Color features involve the use of HSV (Hue-Saturation-Value) colorspace over RGB (Red-Green-Blue) as HSV is more closely related to how humans perceive colors. H is for the colors used, S is for the saturation of colors and V is the lightness of the image. Classification done using Naive Bayes, KNN with 1 neighbour, KNN with 10 neighbours, SVM and ANN. Accuracy of each type of classification and feature extraction compared.

3. From Impressionism To Expressionism: Automatically Identifying Van Gogh’s Paintings[28]

Curators, art historians, and connoisseurs are often interested in determining the authorship of paintings. The automatic identification of Vincent van Gogh’s paintings using a Convolutional Neural Network that extracts discriminative visual patterns of a painter directly from images, and a machine learning classifier for decision process. Divide each painting into non-overlapping patches, classify them, and get the response. In this study, proposed a new and public painting dataset for artist identification, which improves on previous works that did not disclose the data used. It focuses on the classification of paintings into van Gogh and non van Gogh paintings.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

4. Using Machine Learning for Identification of Art Paintings[29]

In this paper a combination of various computer vision features (14) and were able to achieve 85.13 % accuracy in identifying images from a pool of seven artists. By comparing the performance of different features, they found out that HOG2x2 has the best performance in general since it can capture more balanced key points over the images. Finally, that the performances of the different features might imply the preferences and styles of the artists.

5. Feature Selection for Paintings Classification by Optimal Tree Pruning[30]

This work uses a decision tree as classification in order to assess the authenticity of art work. The reason behind using this form of classification is to understand the reasoning and the point of view or perspective of the art expert. A simple data mining tool is preferred that is optimally pruned decision tree in favor of complex tools that involve large feature sets to be extracted from images as the logic is simpler to understand. Smaller feature sets lead to minor errors in classification which gives an ease of understanding and interpreting the results. The database consists of 147 Dijkstra and 160 Wiegers paintings. Decision trees were used to better understand the results and underlying logic of classification.

6. Summary of Literature Survey

Summary

No.	Classes	Images	Features	Algorithm	Highest Accuracy
1	7	14	Hidden Markov Trees	SVM	78%
2	6	70	3	ANN	64%
3	2	331	None	CNN (VGG)	92.5%
4	7	1400	14	Linear SVM	85.13%
5	2	307	3	Decision Trees	72.1%
6		70	4	k-Nearest Neighbor, Hierarchi-cal Clustering, Self Organizing Maps, Multi-Dimensional	Undefined

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

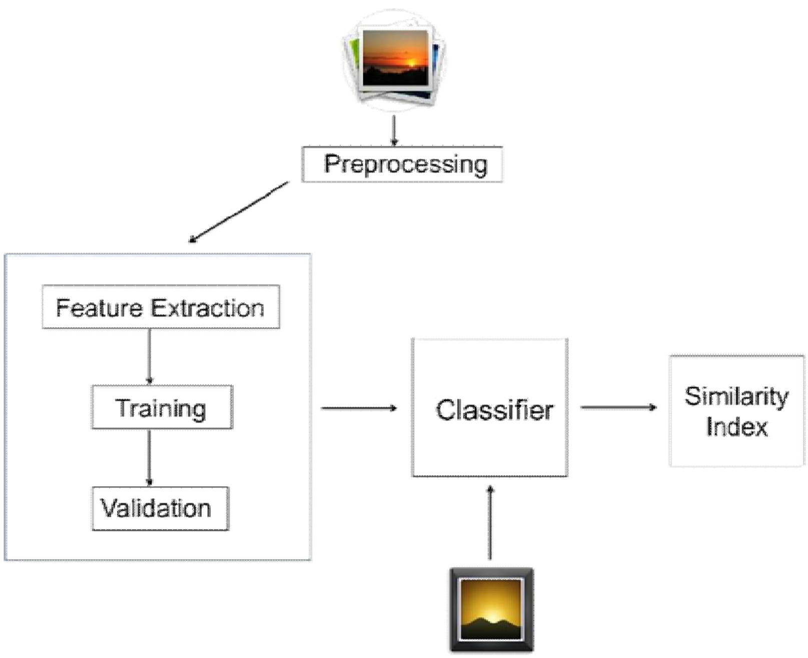
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

				Scaling	
7	19	1633	3	Self Organizing Maps	0.9 Precision
8	2	9	Wavelet Decompo-sition	Statistical Model	100%
9	5	358	3	ANN	55%



ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

IV. PROPOSED SYSTEM ARCHITECTURE

1.System Flow

i. Input training images, which may differ in quality and size are given to the pre-processor

ii. Pre-processing - resizing, scaling, rotating, shearing - takes place to bring the images to the required size (256x256).

iii. Organize the pre-processed images into directories.

iv. Identify the classifier. If CNN, go to step 8 else go to step 7.

v. Performing feature extraction.

vi. Do validation using the validation set images .

vii. An image is given to test the classifier

viii. The image is classified into the classes the result is analysed to calculate the similarity index.

2.Technologies Used

We have used the following technologies for the implementation of our model :

1. Feature Extraction

•OpenCV: It is a computer vision package, implemented using Python 2.7

2. Machine Learning

•Scikit: It is a Python package which is used to implement pre-processing, dimensionality reduction, clustering, classification and machine learning algorithms [18]

•Keras: It is a Python package used for neural networks implementation that can implement both convolutional networks and recurrent networks individually or combined[15].

All these technologies have been used in order to de ne our classifier. We have used 256 x 256 images of a large number, so our system requires a 32 GB RAM to compute the model.

V. FEATURE EXTRACTION

To determine the style of a painter it is necessary to find the unique element in all his paintings.To find what elements can be used to identify a painting that are distinctive to a painting the step of feature extraction is carried out. The feature extraction involves generating feature descriptors that are used to represent an image or a patch of an image by which essential information is extracted and rest unnecessary information is thrown out.While implementing this project, we carried out analysis by using various features, however by trial and error and their application we restricted our feature list to the four features mentioned below:

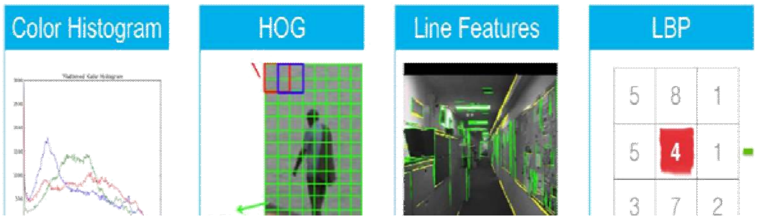
ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017



1. Histogram of Oriented Gradients (HOG)

HOG is a newly found descriptor which is rapidly replacing haar cascade descriptors. This descriptor is employed in the feature extraction phase of images. HOG features aren't that simple. They involve calculation of gradient, object binning and histogram generation of features from blocks.A histogram is a frequency chart. In HOG descriptors the final output is a histogram of the gradients. Gradients can be used to depict changes in intensity of the pixels.Since the edges and corners are generally represented by sharp changes in intensity, gradients play an important role in feature extraction as crucial in-formation is included there.Since images are in fact a vector of pixel intensity values, for each pixel in an image both horizontal and vertical gradients are calculated.Thus for a particular pixel the horizontal gradient is calculated by subtracting the intensity values of the pixels of the immediate left and right neighbouring pixels and the vertical gradient is calculated by subtracting the values of intensities of the pixels above and below it.Kernels [1 0 -1] and [1 0 -1]T when applied to a pixel give the appropriate gradients. Through the horizontal gradient it is possible to detect vertical sharp changes in intensities and through the vertical gradient it is possible to detect sharp horizontal changes in intensities. If an image has RGB color channels then gradients are computed for the three channels separately and the maximum gradient magnitude and out of the three is assigned to that pixel along with the angle. Thus smooth regions are not detected through HOG. HOG is a complete or dense type of feature. It calculates the gradients for all the pixel values and divides the image into blocks say of 16 X 16 or any other size. Thus out of the pixels in the block a particular gradient outcome in terms ofmagnitude and direction is decided by popular vote from the pixels within that block. Normalization of the histograms is carried out for compatibility with the various functions.

2. Color Histograms

Histogram represents a frequency chart. An image generally has color channels like RGB, HSV, CYK. Getting the count and thus frequency of pixels of different colors helps us gain important information about an image. In terms of this project it is possible to identify the colors a painter likes to use that is the color scheme he would most probably utilize in his paintings. Thus a color histogram provides the distribution of colors in images. It may also refer to the count of pixels of a particular color in the image. In case there is a huge wide range of colors then a counting of pixels ranges is carried out.The color space is segmented into different ranges such that the color values in a range have close by intensity values. For construction of a histogram, the color intensities have to be identified in an image then appropriate ranges through which bins are decided and then counting of pixels is carried out. Suppose an image uses the RGB color scheme then the next step that is done is normalization. This done by dividing the pixel intensity value by the sum of total RGB values in the image. Then according to the number of bins counting is carried out. We cannot understand aspects such as details regarding the orientation, shape, texture by analyzing histograms of color. Color histograms are also largely dependent and affected by noise and lighting variances .Also the curse of dimensionality affects the number of bins. This feature is extremely simple and thus easy and quick to compute. To identify known elements or locations color histograms are the best choice. Two dimensional color histograms are an effective way of classification that is new nowadays.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

3. Line features

Lines in an image are different from edges or corners. Line feature extraction involves multiple processes. Parallel, horizontal, vertical and diagonal lines all are of great importance.Steps are :

- Blurring of the image using a filter
- Detecting edges to get binary image
- Using Hough Transform detect lines

Hough transforms are efficient in a way that even if the lines in an image are a bit distorted they have a high probability of getting identified. In the Hough transform a line is represented in polar form. For any pixel in an image using various angles a value corresponding to a Hough transform is stored. Then all pixels having the same value are joined together to represent a
line.

4. Local Binary Patterns (LBP)

Local Binary Patterns (LBP) texture descriptors that are used to find the texture of the image. This is helpful to identify the brush strokes of the artist, that helps in defining his or her style. LBP computes a local representation of the texture instead of global like Haralick texture features. This local representation is calculated in the following steps :

- First, the image is converted to grayscale.
- Next, for each pixel in the grayscale image, a value r is chosen as the neighbourhood of that pixel.
- Then the LBP value for each pixel is calculated by considering the center pixel as a threshold.
- The neighbouring pixels are compared to the center pixel and if are greater than equal to, value is 1, else 0.
- In this fashion a binary number is obtained by considering consistently a start to end pixel. This is the LBP value of the center pixel.
- Now, the center pixel's value is changed to its LBP value.
- All the LBP values of the pixels are calculated in this manner until a grayscale image is obtained as output which shows the texture of the image.

LBP can be scaled to the system’s requirements by deciding the neighbourhood radius r and the number of points p in the neighbourhood which are to be considered to calculate the LBP value.

5. Implementation of Features Extracted

For feature extraction we have utilized OpenCV and Scikit toolkit available in python.For each image we have extracted four features namely :

- Color Histograms to help understand a painters most used color scheme
- Local Binary Patterns to help understand the texture and thus the artist's brush strokes
- Histogram of Oriented Gradients to separate the foreground and background in a painting
- Lines (straight) to figure out the type of painting(e.g. portrait or landscape)

The feature vector and the name are stored in different lists. These lists are used by the training program.

VI. MACHINE LEARNING

Machine learning is a domain of artificial intelligence in which computers do not have to be directly programmed. The computer in fact learns on its own.Machine learning involves identification of patterns and taking actions based on those patterns.In our project we have used both supervised and unsupervised methods of classification.
We have carried out classification using 5 machine learning algorithms - kNN, SVM, clustering, CNN and SVM+kNN:

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

1. kNN :

Since kNN (k Nearest Neighbour)[14] is an instance based algorithm we decided to use as it is a simple and quick algorithm to implement. Since no explicit training is required for kNN we evaluated the model taking different values of k to find that k = 19 gave highest validation accuracy for the model. The features extracted from images are stored in .arff file. After training model with maximum accuracy is selected. Then normalization is carried out to reduce noise which may result in errors. The model has been saved in a binary file formed through features stored in a .arff file which is given as input while testing the model. The training code has been written in python using libraries such a matplotlib, scipy and sklearn.

2. SVM :

For SVM(Support Vector Machine) classifier [22], we have trained a multi-class probabilistic SVM classifier. It is an linear one vs all classifier. For training, features stored in .arff file are used. To find optimum value of C(Cost of constraints violation) We trained SVM with different values of C. After that SVM classifier with maximum accuracy is selected. Normalization is performed on this classifier to avoid error due to noise. The model has been saved in a binary file formed through features stored in a .arff file which is given as input while testing the model. The training code has been written in python using libraries such a matplotlib, scipy and sklearn.

3. k Means Clustering :

We have performed k means clustering on images from twenty five artists. Thus k has been chosen to be 25. For validation we have written a python program that compiles the predicted class, actual class and image name into one .csv file.

4. Ensemble Algorithm :

Ensemble algorithm combines different machine learning models to improve accuracy. We have implemented ensemble algorithm using kNN and SVM. This algorithm selects the maximum similarity of given class with any image from all the results.

5. CNN :

Like every CNN model, our model also has an input layer, few hidden layers and an output layer.This model has been implemented in python using keras and tensorflow libraries. The input layer is of size 256 X 256 X 3 which represents the dimensions of an image from the training data.A batch size of 32 is used which decides the number of images that are processed at one time for training in an epoch.Accuracy is measured in the terms of root mean squared error.The type of the model is VGG COVNET and uses the model type of categorical crossentropy. Architecture of our CNN is as follows: There is one input and one output layer. Number of hidden layers in our model are 2 and it has one penultimate layer which has 1024 neurons. These layers have used filters of size 32, 64, 128 and 256 respectively. There are 3 maxpooling layers one after input layer and one after each hidden layer. The pool size for these layers is 2X2. Each layer has kernel size 3X3. For regularization softmax function is used. 0.5 dropout rate is used to avoid overfitting. The model is trained using 43 epochs. For 2 classes, we have achieved 100 percent training accuracy and 97 percent validation accuracy. For 5 classes, we have achieved 85 percent training and 80 percent validation accuracy.The output layer has 5 neurons since CNN has been tried out for five classes.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

VII. RESULTS

No. of classes	SVM	kNN	CNN	SVM+kNN
2	100	100	100	100
5	70	80	80	70

Table 2: Accuracy of classifiers with number of classes

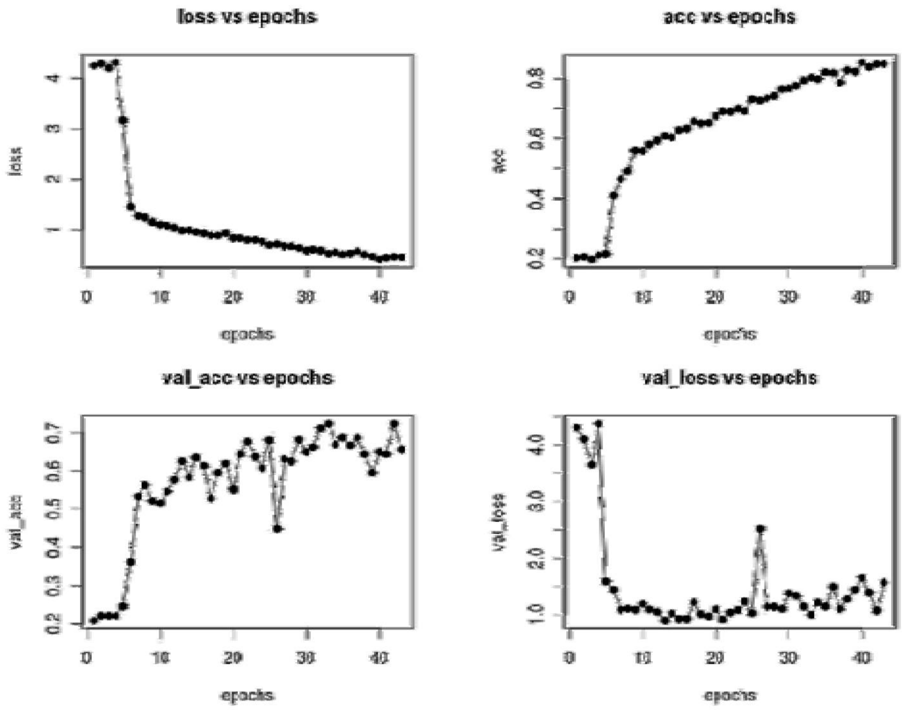


Figure 3: Results of CNN with 1024 neurons, 3 hidden layers and 43 epochs

VIII. ANALYSIS

1. SVM :

The accuracy for SVM is 100 percent when two classes are considered. This is because SVM is a binary classifier. For five classes accuracy is 70 percent. We have used the linear kernel of SVM for classification that uses one-vs-all technique for boundary decisions. Hence accuracy decreases for more classes.

2. kNN :

For kNN classifier we have done a trial and error basis to find out k. With 2 classes, we have achieved an accuracy of 100 percent where k = 5. For 5 classes the accuracy is 80 percent with k = 19. Hence similar to SVM, with more classes the performance decreases, and the number of neighbours (k) increases.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

3. CNN :

We found that the accuracy of our CNN model decreased greatly as the number of classes increased. This is because the number of scanned paintings in these classes vary greatly. Also, the output of CNN depends upon the number of neurons, hidden layers and the epochs.

4. k-Means Clustering :

While performing unsupervised learning on our data, we found out that clusters are formed based on color scheme and not artists as was essential for unknown artists to be classified. Hence through clustering we can see how similar a new artist's color scheme is to the artists in our database. Still we have achieved an accuracy of 25 percent.

5. SVM+kNN :

This is an ensemble technique. It combines both SVM and kNN algorithms to adopt an ensemble technique. We have achieved an accuracy of 100 percent for two classes and 70 percent accuracy for five classes.

6. Individual features :

We took one feature at a time to check which feature individually classifies the paintings best using kNN. After running the code on a single feature we found that accuracy was achieved in a decreasing order of the features given below:

- Color
- LBP
- HOG
- Line

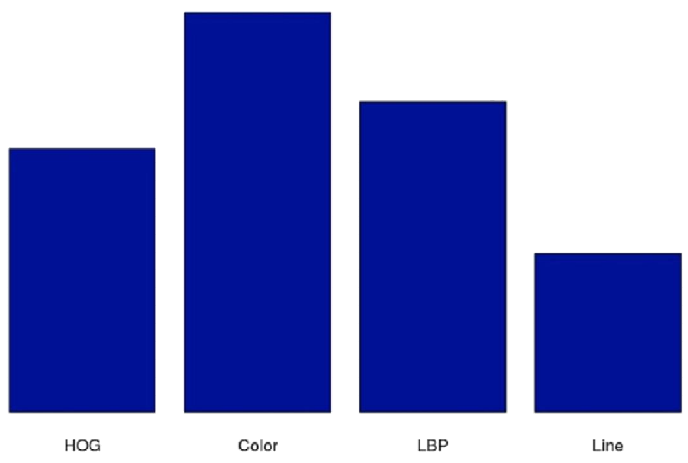


Figure 4: Accuracy with individual features

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

IX. CONCLUSION

Detection of forgery in art paintings is a problem of grave importance in todays world.We have attempted to solve this problem using machine learning techniques both supervised and unsupervised.Our application is not full proof as it does not take in consideration aspects of the painting like age and also is limited to famous artists.
Thus through the application we have completed the objectives of identifying the style of an artists and thus being able to identify an artist from a painting. A user interface has been created which facilitates a user to upload an image of a painting and get back the similarity index with painters whose paintings are used for training.

In the application we have tried out 3 different supervised machine learning algorithms using 2000 number of training images by five famous artists and four features.

We have found that kNN algorithm gives maximum accuracy of 80% on test data of 10 images. CNN has shown maximum validation accuracy on two classes. Also individual features were tested using kNN and the feature color stood out to give highest accuracy.

k-means clustering algorithm was implemented as well. It lead to poor accuracy. The images were in fact grouped into clusters according to a single feature that is color scheme which resulted in poor accuracy.

An ensemble technique was also implemented which combined SVM and kNN to achieve an accuracy of 70%.

REFERENCES

[1] G. Derefeldt, S. Nyberg, J. Alfredson, and H. Allberg, “Is wolfflin’s system for characterizing art possible to validate by methods used in cognitive-based image-retrieval (cbir)?” in *Proceedings of SPIE-The International Society for Optical Engineering, vol. 6057. International Society for Optical Engineering*, 2006, p. 605710.
[2] H. Wolfflin, *Principles of Art History. The Problem of the Development of Style in Later Art*. New York, USA: Dover Publication, Inc., 1932.
[3] P. Kammerer, M. Lettner, E. Zolda, and R. Sablatnig, “Identification of drawing tools by classification of textural and boundary features of strokes,” *Pattern Recognition Letters*, vol. 28, no. 6, pp. 710 – 718, 2007.
[4] R. Sablatnig, P. Kammerer, and E. Zolda, “Hierarchical classification of paintings using face- and brush stroke models,” in *in Proc. of 14th Int. Conf. on Pattern Recognition*, 1998, pp. 172–174.
[5] F. S. Abas and K. Martinez, “Classification of painting cracks for content-based analysis,” *in Machine Vision Applications in Industrial Inspection XI*, M. A. Hunt and J. R. Price, Eds., vol. 5011, no. 1. SPIE, 2003, pp. 149–160.
[6] J. Li and J. Wang, “Studying digital imagery of ancient paintings by mixtures of stochastic models,” *Image Processing, IEEE Transactions on*, vol. 13, no. 3, pp. 340–353, March 2004.
[7] G. Lu, Z. Gao, D. Qin, X. Zhao, and M. Liu, “Content-based identifying and classifying traditional chinese painting images,” in *Image and Signal Processing, 2008. CISP ’08. Congress on*, vol. 4, May 2008, pp. 570–574.
[8] A. Deac, J. van der Lubbe, and E. Backer, “Feature selection for paintings classification by optimal tree pruning,” in *Multimedia Content Representation, Classification and Security*, 2006, pp. 354–361.
[9] T. Lombardi, S.-H. Cha, and C. Tappert, “A graphical user interface for a fine-art painting image retrieval system,” in MIR ’04: *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. ACM, 2004, pp. 107–112.
[10] B. Gunsel, S. Sariel, and O. Icoglu, “Content-based access to art paintings,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2, Sept. 2005, pp. 558–561.
[11] Mosteller, F. and Wallace, D. L. (1984) *Applied Bayesian and Classical Inference: The Case of the Federalist Papers* (Springer, New York).
[12] https://www.svm-tutorial.com/
[13] https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/
[14] https://www.analyticsvidhya.com/blog/2014/10/introduction-k-neighbours-algorithm-clustering/
[15] http://machinelearningmastery.com/tutorial-first-neural-network-python-keras/
[16] http://neuralnetworksanddeeplearning.com/chap6.html
[17] Chapelle, Olivier, Patrick Haffner, and Vladimir N. Vapnik. "Support vector machines for histogram-based image classification." *IEEE transactions on Neural Networks* 10.5 (1999): 1055-1064.
[18] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*12.Oct (2011): 2825-2830.
[19] Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., &Jackel, L.D. (1989). BACKPROPAGATION applied to handwritten zip code recognition. *Neural Computation*, 1(4).
[20] Cottrell, G. W. (1990). Extracting features from faces using compression networks: Face, identity, emotion, and gender recognition using holons. In *D. Touretzky (Ed.), Connection Models: Proceedings of the 1990 Summer School*. San Mateo, CA: Morgan Kaufmann.
[21] Seldin, Y., Starik, S., Werman, M. (2003). Unsupervised Clustering of Images using their Joint Segmentation in *Electronic Imaging*, 2004.
[22] Cusano, Claudio, Gianluigi Ciocca, and Raimondo Schettini. "Image annotation using SVM." *International Society for Optics and Photonics*, 2003.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

[23] D.Zhang et al., Content-based Image Retrieval Using Gabor Texture Feature, *Pattern Recognition*, Volume 45, Issue 1, January 2012, Pages 346–362
[24] C. R. Johnson, Proc. 1st int. *Workshop on image processing for artist identification*, 2007.
[25] S. Lyu, D. Rockmore, H. Farid, "A digital technique for art authentication", *PNAS*, vol. 101, no. 49, pp. 17006-10, 2004.
[26] Polatkan, Güngör, et al. "Detection of forgery in paintings using supervised learning." *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009.
[27] Zujovic, Jana, et al. "Classifying paintings by artistic genre: An analysis of features & classifiers." *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*. IEEE, 2009.
[28] Folego, Guilherme, Otavio Gomes, and Anderson Rocha. "From impressionism to expressionism: Automatically identifying van Gogh's paintings." *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016.
[29] Blessing, Alexander, and Kai Wen. "Using machine learning for identification of art paintings."*Technical report (2010)*.
[30] Deac, Ana Ioana, Jan van der Lubbe, and Eric Backer. "Feature selection for paintings classification by optimal tree pruning." *International Workshop on Multimedia Content Representation, Classification and Security*. Springer Berlin Heidelberg, 2006.