Final Project. Advanced Topics in Deep Learning

Zhenghui Chen, Lei Kang
Department of Chemical and Biological Engineering
HKUST
{zchenef, lkangaa} @connect.ust.hk

Abstract

In this project, we perform data augmentation method by cropping the images of Raphael's paintings and extract the feature vectors by pre-trained deep neural networks VGG16. Then we use transfer learning and other machine learning methods to predict the disputed paintings in Raphael's paintings. We compare the performance of different models and explain the reasons.

1 Introduction

The identification of genuine paintings and forgery paintings is important in art authentication. There are many art authentication methods, such as the judgement from the expert and physical examination. Recently, the authentication process based on image classification has worked well. Liu et al. proposed a geometric tight frame for art authentication of van Gogh paintings [1]. They use feature extraction and outlier detection method to achieve high classification accuracy.

In this project, we attempt to predict the identity of 7 disputed paintings. These paintings may be Raphael's paintings or forgeries. We use two kinds of methods to classify the painting images. Firstly, we use transfer learning method to make the classification. Then we use some traditional supervised learning methods to build the classifiers based on the image feature vectors extracted by pre-trained VGG16. The traditional supervised learning methods include Support Vector Machine (SVM), K Nearest Neighbours (KNN), Random Forest (RF) and Logistic Regression (LR). We tested all the models on the testing dataset and got high testing accuracy above 90% for each method. In addition, we make comparisons among different supervised learning methods and find SVM has the best performance. We give the reasons why some models are better than others based on the bias and variance trade-off theory. Finally, we apply the well tested models to the disputed images to make the prediction whether they are true paintings. We give the probabilities of each disputed image to be a true painting.

2 Dataset

The Raphael's paintings dataset contains 28 images with the format of TIFF, TIF, JPG. In these scanning images of paintings, 12 images are Raphael, 9 images are not Raphael, and 7 images are disputed. The images of TIFF and TIF format contains RGBA channels, we convert all the pictures to JPG format with RGB channels.

As the size of each image is quite large, we crop the paintings into many small patches with the size of 224×224. Therefore, we will make full use of the information from these paintings. For known images, we obtain 4166 images of Raphael paintings and 2938 images of not Raphael paintings. We label the known Raphael image as 1 and not Raphael image as 0. These 7104 images and labels can be used for training and testing.

The whole dataset was divided into two parts, 5000 images for training dataset and 2104 images for testing dataset. Table 1 shows the distribution of dataset.

Table 1: The distribution of dataset

Dataset	Number
training	5000
testing	2104

For disputed images, each image is also cropped into small images for classification. The detail is shown in Table 2. We can predict the identity of all small images and then calculate the ratio to analyse whether the disputed images are Raphael's paintings or not.

Table 2: The dataset of 7 disputed images

Image ID	1	7	10	20	23	25	26
Number	210	493	456	675	80	70	196

3 Feature extraction

3.1 VGG16

VGG was proposed by the Visual Geometry Group of Oxford [2]. It is proved that increasing network depth can improve the accuracy of the network to some extent. VGG is a deeper convolutional neural network and use 3 × 3 convolution filter. It achieves great classification results by pushing the depth to 16–19 weight layers. The pretrained VGG model contains the feature extraction and classification part by using the convolutional layers and fully connected layers.

Figure 1 shows that the structure diagram of VGG16. The VGG16 network includes 13 convolutional layers and 3 fully connected (FC) layers. The input of VGG16 is a fixed-size 224 × 224 RGB image and the output of last FC layer is 1000 channels. In this experiment, we use the pretrained deep neural networks VGG16 to extract the feature vector of the Raphael's paintings. We choose the output of layer fc7 as the feature vector. After feeding the cropped Raphael's paintings image [224, 224, 3] (height, width, channels) to the VGG16 net, we can get the feature vector with 4096 variables. Then we use this feature vector to do image classification.

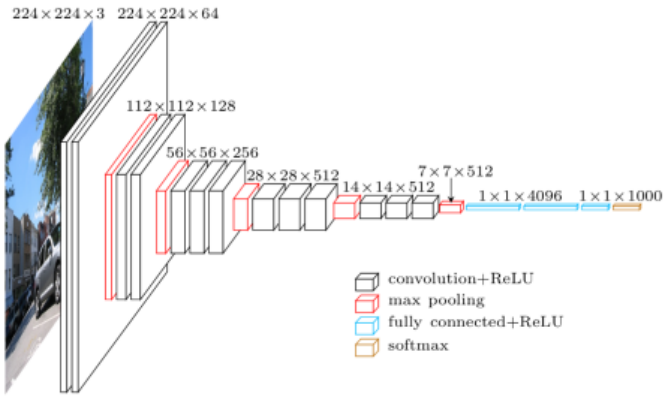

Figure 1: The structure diagram of VGG16

4 Image classification

4.1 Transfer Learning

In transfer learning, we fine-tune the pretrained VGG16 model and do image classification directly on the dataset. We change the last fully connected (FC) layer of the pretrained VGG16 model. The original output of this layer is 1000-dimensional vector as it is designed for the classification of 1000 categories. We change the output of the last FC layer into 2 classes in our experiment.

During training phase, the pretrained weights for the feature extraction part (the convolutional layers) is fixed, we only train the classification part and update the weights of FC layers. After 5 epochs, the model achieves 99% accuracy in the training set and 96% accuracy in the testing set. The results are shown in Table 3.

Table 3: Training results based on transfer learning

Training accuracy (%)	99
Testing accuracy (%)	96

Then we use this model to predict the disputed images. Table 4 shows the prediction results of 7 disputed images. As the disputed image is cropped into many small images and each small image has its own prediction, we calculate the ratio of Raphael images in all small images and use this to measure the possibility of Raphael paintings. For example, Image 23 is most likely to be Raphael paintings because the ratio is high as 84%. Image 20 is not Raphael paintings because the ratio is very low as 3.4%. Image 7,25 and 26 are more likely to be Raphael paintings as the possibility is higher than 60%. Image 1 and 10 is not sure because the possibility is around 50%. These two images may be Raphael paintings or not.

Table 4: Prediction of disputed images based on transfer learning

Image ID	1	7	10	20	23	25	26
Not Raphael	101	180	208	652	13	27	51
Raphael	109	313	248	23	67	43	145
Possibility (%)	51.9	63.5	54.4	3.4	83.8	61.4	74.0

4.2 Traditional supervised Learning method

In addition with the deep learning methods, we also tried some traditional supervised machine learning method to make the classification based on the extracted feature vector. The feature vector extracted by VGG16 has 4096 dimensions, which means there will be a lot of redundant dimensions for the classification task. In order to improve the efficiency, we first use principal component analysis (PCA) method to reduce the dimension of the feature vector. We preserve 100 principal components from 4096 features and retain 97% of feature information. It can be seen from the explained variance graph.

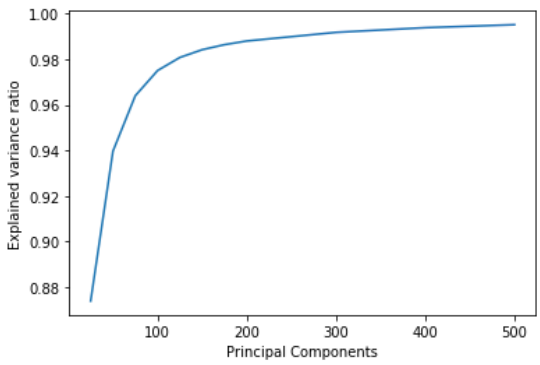

Figure 2: The explained variance ratio of different principal components

4.2.1 KNN

The k-nearest neighbours (KNN) algorithm is an easy method in classification problems. This approach is easy to understand, implement and requires no training process. It determines the category of the sample according to the class of its nearest neighbour.

In order to find the best parameter k for our classification problem, 10-fold cross validation is performed on the training data. The results are shown in Table 3. We can see the accuracy is highest as 98% when the number of nearest neighbour is 3. Then we use this parameter to build KNN model on the whole training dataset and test it on the testing dataset. We get a testing accuracy of 97.4%. Finally the model was used to predict the disputed images.

Table 5: 10-fold cross validation of different nearest neighbour numbers

Nearest Neighbour	1	3	5	7	9	11	13	15
Accuracy (%)	97.5	98.0	97.9	97.7	97.6	97.6	97.4	97.3

Table 6 shows the prediction results of 7 disputed images. Image 20 is not Raphael paintings because the ratio is very low as 4.4%. Image 1 is also not Raphael paintings as the possibility is only 21.4%. Image 7 and 26 is most likely to be Raphael paintings as the possibility is higher than 70%. Image 10, 23 and 25 are not sure because the possibility is around 50%-60%. These three images may be not Raphael paintings.

Table 6: Prediction of disputed images based on KNN model

Image ID	1	7	10	20	23	25	26
Not Raphael	165	133	220	645	29	26	32
Raphael	45	360	236	30	51	44	164
Possibility (%)	21.4	73.0	51.8	4.44	63.8	62.9	83.7

4.2.2 Logistic Regression

Logistic Regression (LR) is a generalized linear regression analysis model. Sigmoid function is used to strengthen nonlinear factors and deal with binary classification problem easily. In the experiment, we train the LR model based on training dataset and test it on the testing dataset. The LR model achieves the testing accuracy about 93.9%.

The prediction results of disputed images are shown in Table 7. Image 23 and 25 are Raphael paintings, because their possibilities are higher than 80%. Image 7 and 10 is also likely to be Raphael paintings with the possibility close to 70%. Image 20 is not Raphael paintings as the ratio is lower than 4%. Image 1 and 26 are possible to be Raphael paintings as the possibility around 60%. They may be not Raphael paintings.

Table 7: Prediction of disputed images based on LR model

Image ID	1	7	10	20	23	25	26
Not Raphael	94	143	120	651	10	11	71
Raphael	116	350	336	24	70	59	125
Possibility (%)	55.2	71.0	73.7	3.6	87.5	84.3	63.8

4.2.3 Support Vector Machine (SVM)

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary classifier. SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. Here we use the radial basis function kernel to make the classification.

For the model fitting process, we randomly divide the 7000 image vectors into two datasets, training dataset (5000 image vectors) and testing dataset (2000 images vectors), then we build SVM model on the training dataset and test it on testing dataset. The accuracy on the testing dataset is 98.2%.

The prediction results of disputed image patches are shown in Table 8. Image 7 is most likely to be Raphael paintings and the possibility is higher than 90%. Image 23, 25and 26 is also likely to be Raphael paintings with the possibility around 70-80%. Image 20 is not Raphael paintings as the ratio is lower than 5%. Image 1 is also not Raphael paintings as the ratio is lower than 40%. Image 10 may be not Raphael paintings because the possibility is only about 50%.

Table 8: Prediction of disputed images based on SVM model

Image ID	1	7	10	20	23	25	26
Not Raphael	134	28	209	642	18	20	34
Raphael	76	465	247	33	62	50	162
Possibility (%)	36.2	94.3	54.2	4.9	77.5	71.4	82.7

4.2.4 Random Forest

Random forests or random decision forests are an ensemble learning method for classification, which operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) of the individual trees [3]. We build the random forest on the training dataset and test it on the testing dataset. The accuracy on the testing dataset is 94.4%.

The prediction results of disputed image patches are shown in Table 9. Image 25 and 26 are most likely to be Raphael paintings and the possibility is higher than 80%. Image 7, 10, 23 are also Raphael paintings with the possibility over 70%. Image 20 is not Raphael paintings as the ratio is lower than 20%. Image 1 may not be Raphael paintings because the possibility is only 55%.

Table 9: Prediction of disputed images based on RF model

Image ID	1	7	10	20	23	25	26
Not Raphael	95	109	126	562	18	11	26
Raphael	115	384	330	113	62	59	170
Possibility (%)	54.8	77.9	72.4	16.7	77.5	84.3	86.7

4.2.5 Comparison of KNN, logistic regression, SVM and Random Forest.

A receiver operating characteristic curve, or ROC curve illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied [4]. Receiver operating characteristic (ROC) curve is one of the most effective evaluation metrics because it visualizes the accuracy of predictions for a whole range of cut-off values. Here we get the ROC curve of four different classifiers.

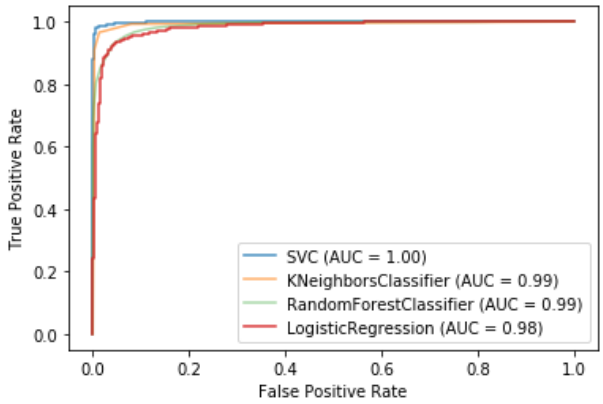

Figure 3: The ROC curve of four different classifiers

Generally, a perfect model will have an AUC (area under the curve) that equals to 1, which means the true positive rate always equals to 1 and the false positive rate always equals to 0. In this situation, the model is able to classify all the samples correctly. So in terms of AUC, the closer is the value to 1, the better is the model performance. Here, SVC has an AUC equals to 1. In other words, SVM has highest true positive rate and lowest false positive rate for this dataset compared with other methods, which represents that it has the best performance. Since all the AUC scores of the four methods is close to 1, it shows that these four models we choose perform well on the image datasets.

Table 10: Testing accuracy for four methods.

Methods	KNN	LR	SVM	RF
Accuracy (%)	97.4	94.0	98.2	94.4

Similarly, the testing accuracy for these four methods also supports that SVM has the best performance among these methods. KNN ranks second and Random forest ranks third. Logistic regression has the lowest performance compared with other three methods, however, since the testing accuracy of four methods is above 90%, this means all these models are suitable for our image classification task.

Why SVM preforms best? This can be explained by the bias and variance trade-off. Among all these four methods, Random forest has the highest variance and lowest bias since it is tree-based method. This means Random forest can fit the complex training data very well, but has lower performance on testing data because of the overfitting problem. In contrast, Logistic Regression (LR) has the lowest variance and highest bias since it's based on the linear combination of predictors. This is the reason why LR can't fit the complex training data well, so it will also not have a very good performance on the testing data. The variance of KNN and SVC is between Random forest and LR, which means KNN and SVM can fit the training data better that LR, but will not over-fit the training data like Random forest. So SVC and KNN will perform better than Random forest and logistic regression. Also considering that KNN is a local fitting method while SVM considers the relationship of all the sample points, SVM can perform better than KNN. So the bias and trade-off can explain why we got such performance rank in terms of ROC and testing accuracy for these four methods.

4.3 Summary of prediction results on disputed images

For the transfer learning method, if we set the possibility threshold to 0.6, then disputed images 1, 10, 20 can be regarded as fake paintings while others are true paintings.

For traditional supervised learning methods, Figure 4 shows the prediction results of 7 disputed images of four methods. Image number from 1 to 7 represents the disputed image 1, 7, 10, 20, 23, 25, 26 respectively. Set 60% as the threshold, image 1, 10 and 20 are not Raphael paintings due to the low possibilities. The other images are likely to be Raphael paintings with the higher possibilities.

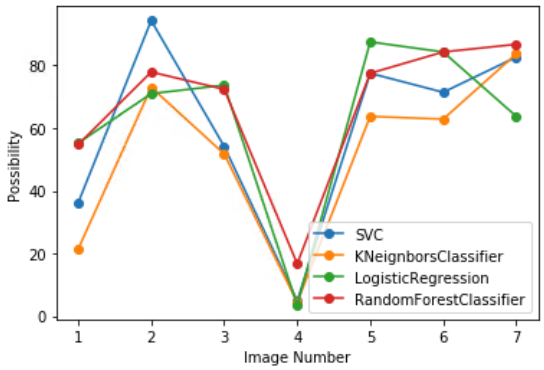

Figure 4: The prediction of disputed images based on four classifiers

The prediction results of transfer learning are in alignment with the results of traditional supervised learning methods.

5 Conclusion

In this project, we use two kinds of methods to classify the painting image patches. First we use transfer learning method to make the classification. Then we use some traditional supervised learning methods to build the classifiers based on the image feature vectors extracted by pre-trained VGG16. We tested the all the models on the testing dataset and got high testing scores which could justify the high performance of the models. In addition, we make comparisons among different supervised learning methods and give the reasons why some models are better than others. Finally, we apply the well tested models to the disputed images to make the prediction whether they are true paintings. We show the probabilities of each disputed image whether it is a true painting. If we use 0.6 as a threshold, disputed images 1, 10, 20 can be regarded as fake painting while others are true paintings.

6 Contribution

Zhenghui Chen: Code for the feature extraction, transfer learning and KNN
Zhenghui Chen: Report writing for the feature extraction, transfer learning and KNN part and make all the tables in the paper. Polishing the whole paper.
Lei Kang: Code for SVM, Random Forest, Logistic Regression
Lei Kang: Write the analysis of SVM, Random Forest, Logistic Regression, and make the comparison of four supervised learning methods. Write the summary and conclusion. Make the PPT and Record the vedio.

References

[1] H. Liu, R. H. Chan, and Y. Yao, “Geometric tight frame based stylometry for art authentication of van Gogh paintings,” *Appl. Comput. Harmon. Anal.*, vol. 41, no. 2, pp. 590–602, 2016.
[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
[3]Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests" *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20 (8): 832–844
[4]Peres, D. J.; Cancelliere, A. (2014-12-08). "Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach". *Hydrol. Earth Syst.* Sci. 18 (12): 4913–4931
[5]Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.