Identification of Raphael’s paintings from the forgeries

Liang Zhicong

[School of Mathematics, Sun Yet-Sen University](http://math.sysu.edu.cn/)

<liangzhc3@mail2.sysu.edu.cn>

Abstract

Given 28 paintings, 11 of which are known as Raphael’s and 10 are known as non-Raphael’s, our task is to predict the authorship of another 7 disputed paintings. This project can be divided into two parts: (1) feature extraction and (2) classification.

For the first part, I applied the geographic tight frame method(Liu,2015). For the second part, I try different methods :(1) Distance Discriminant Analysis(Liu,2015) and (2) Neural Network.

After looking into the misclassified paintings of distance discriminant analysis, I surprisingly found that all the misclassified paintings belong to Raphael. According to this fact, after applying Distance Discriminant Analysis, I applied Neural Network to re-predict the “outliers. If a painting is predicted as outlier but has a probability large than 70% to be Raphael in Neural Network, we still classified it as Raphael’s. We named this method as Combined Model.

In leave-one-out cross validation test, Distance Discriminant Analysis reaches an accuracy of 80%, Neural Network reaches about 92.5% while Combined Model reaches about 96.5%.

Finally, I would like to dig into Logistic Regression.

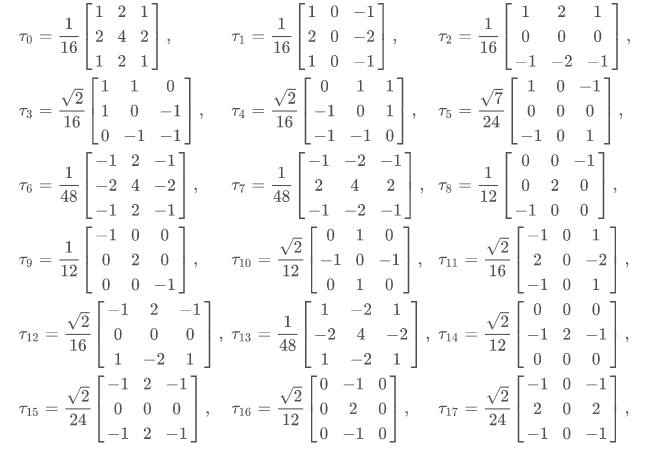
All the codes of this project could be found on my Github repository: [Identification-of-Raphael’s-Paintings](https://github.com/ZhicongLiang/Identification-of-Raphael-Paintings).

Processing Data

Firstly, I transform all the 27 RGB paintings into gray-scale images by and then cut out the edges of them. For some reasons, the 28th painting could not be read in, so it is abandoned.

Feature Extraction

I applied the 18 geometric tight frame filters (Liu, 2015), using a neighbor size of 5\*5:



Then we got 18 feature-images of each painting. We calculate 3 statistics of each feature-image:(1) the mean (2) the standard deviation (std) (3) and the proportion of pixels that are more than one standard deviation away from the mean (tail). So, we have 18\*3=54 features for every painting.

**Distance Discriminant Analysis**

Before distance discriminant analysis, we will first select some most discriminating features. We hope that under these features, Raphael’s paintings concentrate in a circle around their mean while the non-Raphael’s lay as outliers. Mathematically, we define the best features as those with largest AUC.

In experience, I found that normalizing the features or not will give different results. If we normalize the features first, then the best choice would be [6,33,2], which means the mean of , and the tail of . If not normalize, then the best choice would be [18,24], the mean of and .

Main result of **Distance Discriminant Analysis**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Normalized** | **Features** | **Accuracy** | **1st** | **7th** | **10th** | **20th** | **23th** | **25th** | **26th** |
| **🗸** | [6,33,2] | 80% | 🗴 | 🗴 | 🗸 | 🗴 | 🗴 | 🗸 | 🗸 |
| **🗴** | [18,24] | 80% | 🗸 | 🗸 | 🗸 | 🗸 | 🗴 | 🗸 | 🗸 |

Note: 1st, 7th, ……26th represent the 7 disputed paintings, 🗴 means is forgery while 🗸 means its authentic. “Accuracy” is the accuracy of the leave-one-out cross validation test.

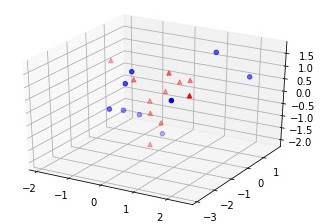
Misclassified Paintings in Cross Validation Test

|  |  |  |  |
| --- | --- | --- | --- |
| **Normalized** | **Features** | | **Misclassification** |
| **🗸** | | [6,33,2] | 2(R), 5(R), 22(R), 27(R) |
| **🗴** | | [18,24] | 4(R), 11(N), 13(N), 19(N) |

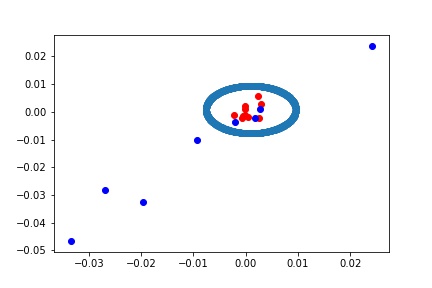
“R” means this painting is Raphael’s while “N” means non-Raphael’s

According to the two tables above, a very important point should be proposed is that all the misclassified paintings in normalized model belong to Raphael, which means this model has a high False Negative Rate(FNR). We can improve it by re-predict those paintings that are classified as outliers and try to lower the FNR, which will be discussed in section “Combined Model”.

Here, we will visualize the two methods. The radius of the circle in the second figure is the best threshold of classification on all the data (instead of leave-one-out cross validation):



Normalized: Red points are Raphael’s while blue ones are Non-Raphael’s.

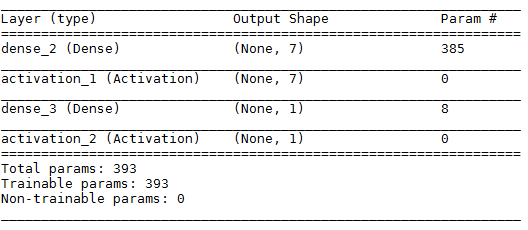


Non-Normalized: Red points are Raphael’s while blue ones are Non-Raphael’s

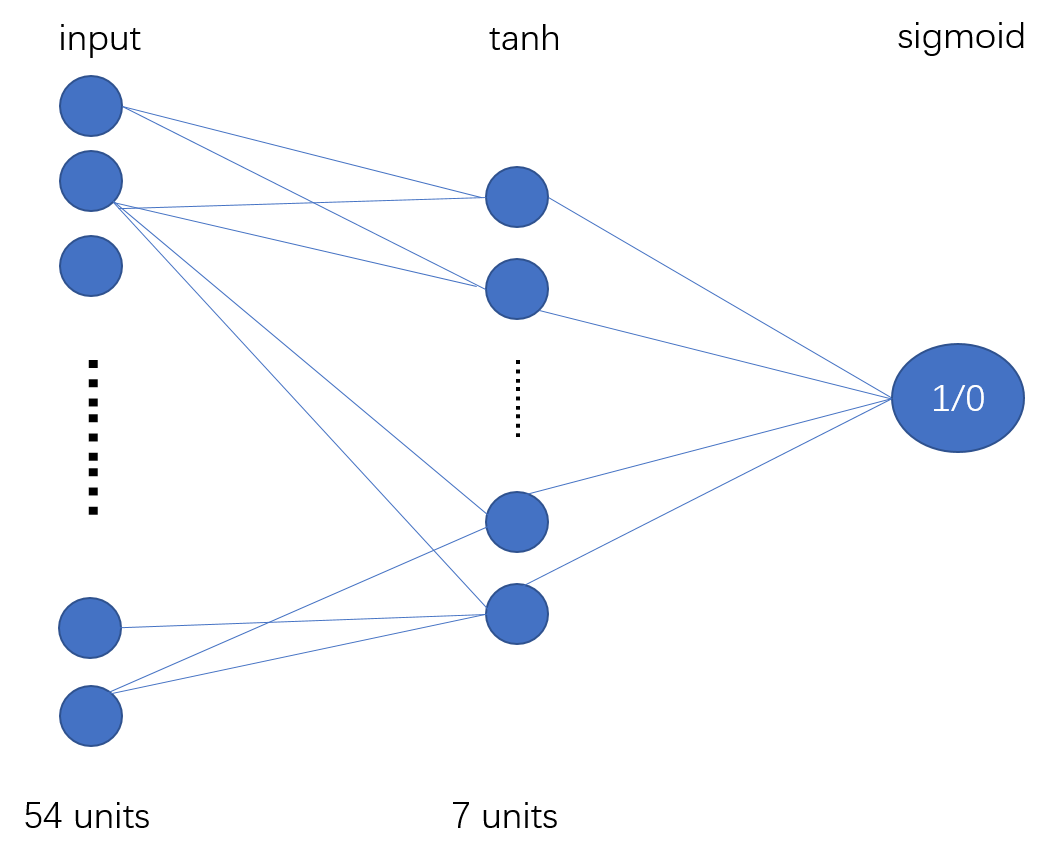
The circle is the classification threshold.

**Neural Network**

I use a neural network, with one hidden layer of 7 units, as my second classification method and reach an accuracy of about 92.5%. Model summary is as follow:



Activation\_1 is tanh and Activation\_2 is sigmoid



**Main result of Neural Network**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **Average Accuracy** | **1st** | **7th** | **10th** | **20th** | **23th** | **25th** | **26th** |
| All 54 features | 92.5% | 🗸 | 🗸 | 🗸 | 🗸 | 🗴 | 🗴 | 🗸 |
| **Average Probability** | | 0.95 | 0.83 | 0.56 | 0.76 | 0.15 | 0.4 | 0.63 |
| **Misclassification** | | 6(R), 13(N) | | | | | | |

R” means this painting is Raphael’s while “N” means non-Raphael’s

**Combined Model**

As we have mentioned above, the normalized Distance Discriminant Analysis model gives a pretty good True Positive Rate (TPR) while a bad False Negative Rate (FNR). In this section, I try to re-predict those paintings that are first considered as non-Raphael’s using the same Neural Network above.

And the classification rule will be:

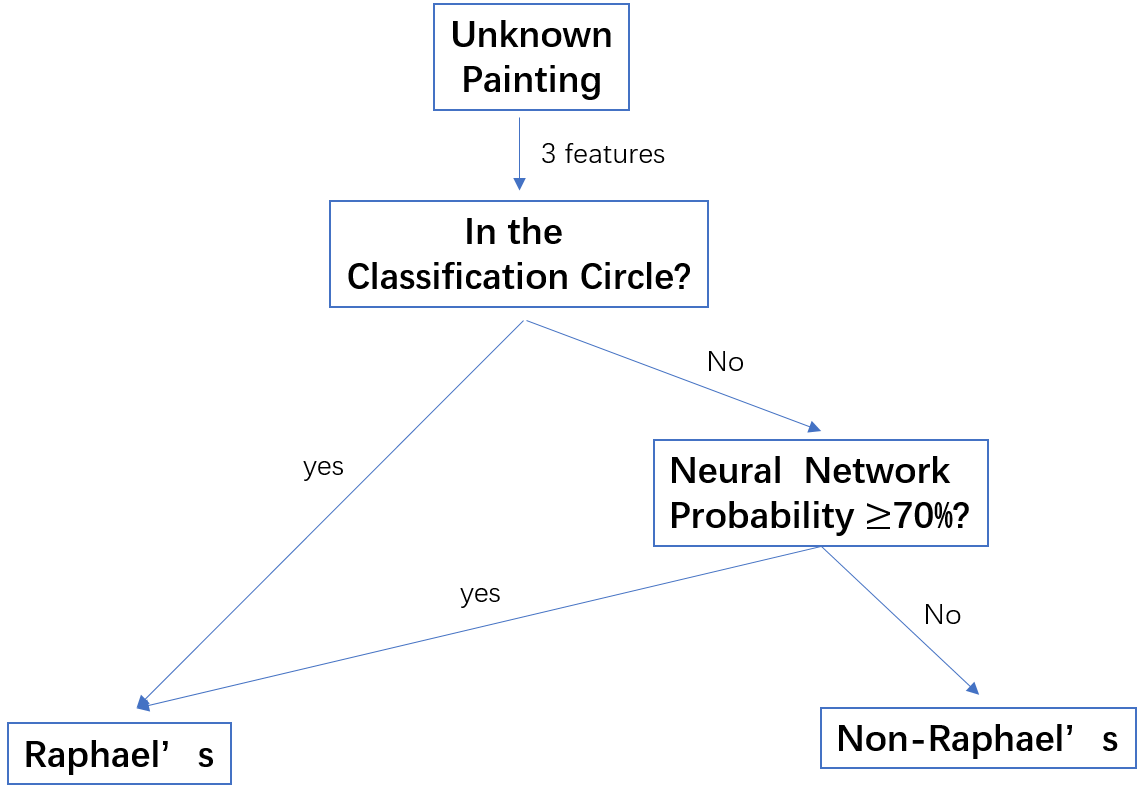
1. If a painting is considered as Raphael’s in Distance Discriminant Analysis, then we accept it as Raphael’s;
2. If a painting is considered as non-Raphael in Distance Discriminant Analysis but has a probability larger than 70% to be Raphael’s in Neural Network, we accept is as Raphael’s;
3. All the other cases will be rejected as Raphael’s.

Main result of **Combined Model**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Accuracy** | **1st** | **7th** | **10th** | **20th** | **23th** | **25th** | **26th** |
| 96.5% | 🗸 | 🗸 | 🗸 | 🗴 | 🗴 | 🗸 | 🗸 |
| Frequency | 0.5333 | 0.7 | 1 | 0.2667 | 0 | 1 | 1 |

“Accuracy” is the accuracy of the leave-one-out cross validation test.

“Frequency” is the frequency of a particular painting classified as Raphael’s under 30 repeated experiments.



Basic Procedure of Combined Model

**Analysis**

In the first model, Distance Discriminant Analysis, we select 3 features under which Raphael’s paintings will concentrated around a specify circle while non-Raphael’s ones will spread as outliers.

However, this hypothesis seems too strong that we may misclassified some Raphael’s paintings as Non-Raphael’s. This case is very likely to happen since we only choose such a small number of features: it is difficult even for Raphael to perform consistently on a few particular features. A re-classification process seems reasonable since even a painting lies as outlier under several features, it is still possible to be Raphael’s if other features resemble those of Raphael’s ones.

In Combined Model, we first classify paintings using the circle hypothesis, then we re-predict the outliers based on a non-circle model, trying to capture as more information as we can.

**Conclusion**

From the result of the two classification methods we discuss above, we are confident that the 7/10/25/26th are Raphael’s , 20/23th are non-Raphael’s while 1th is still under uncertainty.

C:\Users\ZhicongLiang\AppData\Local\Microsoft\Windows\INetCache\Content.Word\1.D.TIF

1th (Uncertain)