MLP Coursework 4 : Project Final Report

Team G22 : s1717961 s1742237

Abstract

Investigating the similarity between images takes a large part in computer vision area. Developing efficient and high prediction accuracy deep neural network can help to deal with such image classification problem, especially with the support from recent advances in deep neural network involving image analysis. Convolutional neural network(CNN), Siamese network and one shot learning approach are explored in our experiments in the hope to better solve such image task through analyzing a subset of Painter by Number dataset. Before this, we also conducted a lot of experiments to adjust parameter settings in basic CNN network in order to make it get optimal performance. Among them, Siamese convolutional neural network with one shot learning perform the best in predicting whether two paintings belong to the same painter or not.

1. Introduction

Determining the authority of the art painting is the most vital part in authenticating these artworks which can directly affect the painting's money value in practice. For example, in art market, savvy collectors try to make investments in emerging and established artists for enormous profits(Anderson, 1974). In this case, forged painting is more difficult to identify since a large amount of forged works circulating in the market. At the same time, identifying the authority of paintings is also one of main concern in art area. Fortunately, recent advances in machine learning especially in realm of image analysis could help people to differentiate the similarity of paintings even though it is still difficult to make such automatic transition and offer convincing judgments. This project select a subset from original data set in an interesting competition, Painter-By-Numbers, hoping to accomplish this painting classification task in the support of computer vision techniques.

As mentioned in our coursework3, painting analysis has been developed with various methods those mostly are based on styles. To simplify this, we proposed to classify art paintings on the basis of artists because we want to know whether random two paintings belong to the same artist or not, instead of judging whether they belong to the similar style. The primary goal of this report is focused on developing an efficient and high accuracy system in predicting the painting's authority regardless of known authors or unknown ones. In addition, in this paper, Our research questions put more emphasis on comparing three models including our baseline system introduced in our previous work, Siamese network and Siamese network with one shot learning because the first two research questions, the impact of hyper-parameter settings and strategies in discouraging over-fitting have been discussed in detailed before.

This report composes of five sections. The following methodology part will describe the basic structure of our three systems and their separate characteristics.Then experiments and result analysis will be explored in the experiment part. Related work discusses a range of studies that have been conducted in relation to our research direction, focusing on our introduced new models.In the end, the conclusion part states the main findings in our study and future interesting research areas.

2. Methodology

In this part, we aims to introduce the methodology used in this project. We will introduce it mianly from three aspects. One is to introduce the baseline architecture we used in this projects. Siamese architecture applied to realize unsupervised learning. and the advanced architecture uses the one-shot learning method which we hope to increase the performance of Siamese.

2.1. Baseline Architecture

Figure 1 describe the convolutional network (with 5 CNN layers) we used in the first project model clearly. We have done several hyperparameter testings and get excellent performance model with five or eight convolutional layers. The first two convolutional layers have 16 filters and the last three have 32 filters in CNN 5 network but with 64 filters in CNN 8 network. The size of every filter is 2*2. We also apply batch normalization after each CNN layer and dropout with a probability of 0.5 to avoid overfitting.

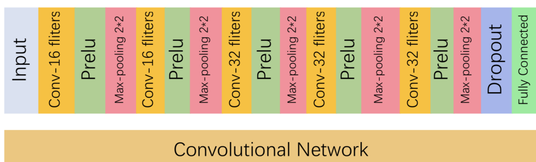


Figure 1. Baseline Architecture

We will continue to use this convolutional network in this project and use it to extract painting features in our new Siamese model. In order to ensure the comparability of the experiments, we will use the same experiment settings with last project. By comparing the prediction performance, we hope to know whether our proposed Siamese and advanced architecture will perform better in judging two paintings belong to the same artist or not.

2.2. Siamese Architecture

However, if we process our task as a classification problem, what we did is just taking advantage of unsupervised learning to finish such classification question. Apparently, its accuracy can be increased, if we change to use supervised learning. As a result, Siamese network is our first choice because of its advantages in deal with similar tasks. For example, siamese model is able to train paintings with any labels even through the label has not appeared in the training set before. Siamese outperforms CNN Network in this respect.

The core thought of Siamese architecture is shown in Figure 4, the network can be basically divided into three parts. The first part is the unique two inputs. The second one refers to the twin networks which share the same set of weights. The aim of the twin networks (CNN here) is still to extract painting features as we discussed above. For the remaining part of the network, it aims to calculate the possibility whether belong to label 1 or 0 on the basis of euclidean distance and difference between prediction value and true label to judge whether the two paintings are from the same artist. The output label of Siamese network is 1 and 0, which stands for the same class and different class separately in terms of the affiliation of two paintings according to the paper of Hadsell et.al(Hadsell et al., 2006).

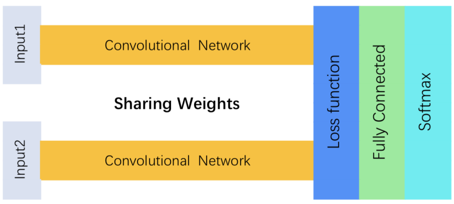


Figure 2. Siamese Architecture

1.Two inputs design
To implement training process, single CNN network needs to use forward propagation and calculate loss value twice. Then, it needs to pass the loss values back to adjust the weight. This process is very cumbersome and requires manual control, so Siamese network is designed as a two inputs network to avoid such troublesome process(Bromley et al., 1994). The two inputs of Siamese network is a pair of images with one label (same 0 or different 1)coming from the twins CNN networks rather a single painting in our original CNN network. With reference to the description in the paper of Chopra et.al(Chopra et al., 2005) and the claim of Norouzi et.al(Norouzi et al., 2012). For example, inputs from this twins network may be a pair of paintings belong to the same artist with label 0 or a pair of paintings belong to different artists with label 1.

2. Loss function
The loss function is defined as the function below.

$$L(F_1, F_2, Y) = \frac{1}{2}(1-Y)D(F_1, F_2)^2 + \frac{1}{2}Y max\{0, m - D(F_1, F_2)\}^2$$

The function we applied is borrowed from the Bromley et.al (Bromley et al., 1994). In the loss function, F1 and F2 represent the two images in a pair. Y means the label of the pair with value 1 or 0. We use Euclidean distance to measure the similarity between two images, which is represented by D(F1,F2) . M is defined as 1 to help us calculate punishment, so we can get a better weight value. This loss function can be understood easily. For a pair of images with label 0(the same artist) the value of similarity is larger while the value of loss function become smaller for a pair of images with label 1(different artists). In other words, the value of similarity is smaller, the value of loss function is lower because their distance is closer.
In comparison with the baseline system, Siamese network solves the "painter by numbers" problem by turning traditional classification method (60 classes for 60 authors) into a simpler judgment, yes (same class) /no (different class). But the primary difference between these two networks is unsupervised learning we adopt before with contrast to supervised learning.

2.3. Advanced Architectures

Based on the first two models we introduced, an advanced architecture is built in the hope to improve prediction accuracy in solving our task. We want to achieve better performance in validation set even with our small data set so that we searched various suitable algorithms in handling limited training data problem and chose one shot learning. In most cases, learning good features with only few examples could be expensive and difficult. One-shot learning method is a powerful method in this respect. That is why we combine this learning with the second model. The method applied in Siamese network could achieve decent performance even only training a small dataset relatively easier, using naturally rank similarity between inputs(Koch et al., 2015). To apply the one-shot method, according to the methods description in the article of Koch.G et.al(Koch et al., 2015),we have a test painting X. There are also some column vector, which we wish to classify into one of C classes(60 in our project). We are also given some other images $xcCc = 1$, a sequence of columns stand for the examples coming from each of the class. We use X and xc as our inputs of the advanced architecture. Then we regarded the predict result belongs to the class which the location of maximum similarity correspond to. The function shows below explain the prediction idea clearly with reference to the claim that one-shot was proposed(Fe-Fei et al., 2003).

$$C = argmax_{c} p(c)$$

Similarly, this complex Siamese network attempts to give a big distance between dissimilar paintings and a small distance between similar paintings. Softmax is used to threshold this label to 1 or 0, then we can get the accuracy and confusion matrix of our model.

3. Experiment

In order to solve our proposed painting classification task, we trained three models on a subset of Painter-By-Number dataset that we described in detail in our last work. Although it is the same dataset, we got different input combinations according to features of distinct models which will be explained separately in the section 4.1. Then we describe how these experiments were built with respect to models in the later part. Result analysis and evaluation are discussed in the end of this experiment section.

3.1. Data

As mentioned in the last work, we continue to take advantage of the balanced data set used in the baseline system. To be specific, our training data contains 60 painters with 100 paintings per painter. And we split this data set with the percentage of 9:1, so that 5400 paintings for training set and 600 paintings for validation set. Although such data set seems to be small, we have to consider the random assortment between two examples($N_{paris} = (6000 * 5999)/2 = 17.997 million$) that is already pretty large for our limited memory. The 60 painters are randomly selected from 125 painters whose owned paintings ranging from 100 to 200 (the second mode in the data set distribution of the competition). For the data preprocessing part, We also resized the dataset into 64*64 pixel cropped from the center of each painting in order to satisfy the fixed-size inputs requirement of CNN network.

Besides, it is necessary to note that we take a very different combination method in order to construct pairwise dataset in the siamese network, despite that paintings and painters datasets are not changed. For example, we classify paintings into 60 classes in the baseline while we divide these pictures into two classes, the same class and different class in the siamese network. For the same class (positive match) we iterate each author in the total 60 and then sample two different paintings that belong to this author, making pairwise combination from the selected painting to 9 paintings after this one(see the below figure where P1-P10 represent the painting id). For the another class, a negative match is carried out, we chose two different writers and sample one painting from one of them then randomly sample one painting that belongs to different painter nine times. Ideally, we hope to explore larger datasets, however, we were not able to try this due to memory and time constraints. In total, we conduct 78720 samples in training set and 1920 samples in validation set with equal numbers of positive and negative examples. The main difference in the siamese combined K shot learning compared to the simple siamese is that the selection of training data numbers per class on the basis of K value.

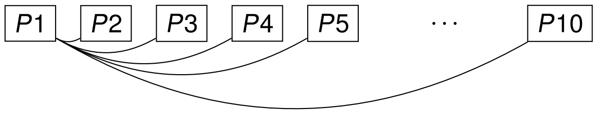


Figure 3. Create Pairs

3.2. Experimentalal Setup

We conduct experiments in the hope to solve problems in identifying whether two paintings belong to the same painter or not, with satisfying accuracy as we expected. In particular, trails are designed to answer the following questions:

- which model can achieve higher validation accuracy.
- What factors may lead to the performance differences.
- whether the best model equips with decent generalization ability as well.

To answer these questions, we evaluate our three models (see section 2) on the same dataset coming from an interesting competition "Painter by numbers" aiming to improve prediction accuracy in revolving around the same basic task. However, each model has different pair combinations, which leads to different input vectors despite with the same paintings. For instance, that is why we used categorical crossentropy to compute our training loss while we adopt a classical loss function, contrastive loss, in our new two models Siamese network and Siamese network with one shot learning. Despite this, both simple Siamese and complex network use the baseline CNN as their basic model structure. And all of models were built using sequential model in Keras library, a popular python deep learning library.

In our previous experiments, we did a lot of trails targeting our first two research questions. Firstly, we explained and tuned our hyperparameter settings in our original CCN model.

1. Activation function, we chose PReLu(He et al., 2015), an improvement of ReLu adding only a small number of parameters.
2. Learning rule(Vincent Dumoulin, 2016), Adam helps to make parameters relatively stable. For example, the learning rate in each iteration within a certain range after correction.
3. Learning rate, ranging from 0.1 to 0.00001 with 10 times decayed as validation accuracy increased, but our line curve is very steep. After trying more rates, we decided to compare 0.0001 and 0.00001 further in the below part.

4. Numbers of CNN layers, we tried 5 CNN layers till 14 layers, increasing with 3 layers, then select 5 and 8 CNN layers according to previous experiment results.

Secondly, various strategies in mitigating likely overfit because of s small dataset are applied and tested to determine suitable values. At the same time, Batch Normalization was added after each CNN layer as well.

a). L2 Regularization, three values were tried, 0, 0.3 and 0.03. 0.03 is picked since their performance is very close.

b). Dropout rate was set to 0.5 after comparing the accuracies of 0.3, 0.5, 0.8 and 1.

Taking memory and time constraints into consideration, we also fix some hyperparameters without extra explanations since it is not the interesting part of the paper:

- The number of training epochs is 200.
- Batch size is fixed as 48.
- The hidden units of every hidden layers is 100.

  In practice, the trained maximum 300 epochs for each network, but monitor their validation error using Early Stopping method. It means that if validation error did not decrease for further 20 epochs, we stopped and save values at this best epoch. Otherwise, we will save the final model status during training procedure as the validation error continued to reduce for the entire learning process.

3.3. Performance Evaluation

Next, we contrast different architectures to determine which one is most successful in predicting paintings' authorities after determining data inputs and basic experiment setting. Since the Siamese networks we introduced are based on two same convolutional networks, like twins, the above settings just like the subset we selected before remain the same when we built new models, for the sake of comparability. However, a few parameters are explored further since we failed to decide their specific values in the last work. The basic structure of our three models has been described in the architecture part in section 2. In this section, we focus on comparing experiment performances of baseline system, basic Siamese network and Siamese network together with K shot learning method on both training and validation set. Then applying the best deep network in test set so as to examine that if it can get expected performance even for unknown paintings with known and unknown painters in the test set.

3.3.1. Parameter Adjustments

To begin, learning rate (0.0001 and 0.00001) and the number of CNN layers (5 and 8) will be investigated again not only because of their close performance in our last report, but also because of their important effect in training cost(Bishop, 2014) and model complexity.

At first, we keep learning rate at 0.0001 when we contrast the impact of 5 CNN layers and 8 CNN layers in deep neural networks. It can be seen from the table 1, as the depth of network increased, we noticed an increase in validation accuracy as expected in the baseline system. In contrast, we note that there was a trend of diminishing returns on validation accuracy in two Siamese network when the number of CNN layers increased from 5 to 8. We suspect that overfit impact on the Siamese network larger than single CNN model due to reduced pairwise data set even though we set regularization method as the same. Since only 2.14% accuracy rise in baseline, we decided to apply 5 CNN layers in our later experiments in order to mitigate overfitting behavior.

After this, we conducted similar experiments except that learning rate become to 0.00001 when 5 CNN layers are used in each model. Although the rise in accuracy is slight, lower learning rate does obtain better performance in all three architecture without obvious extra time consuming. Thus, learning rate is confirmed to be 0.00001.

Model	Learning rate	CNN layer	Acc(valid)
Baseline	0.0001	5	34.50%
Siamese	0.0001	5	65.28%
Siamese&One-shot	0.0001	5	66.33%
Baseline	0.0001	8	36.64%
Siamese	0.0001	8	63.96%
Siamese&One-shot	0.0001	8	65.00%
Baseline	0.00001	5	35.12%
Siamese	0.00001	5	65.53%
Siamese&One-shot	0.00001	5	68.67%

Table 1. Results for different models

In addition, we used the confirmed parameters and other unchanged settings to compare the the impact of K value in K shot learning, thereby determining the best value so as to apply later.

K-way	Err(train)	Err(valid)	Acc(train)	Acc(valid)
10	0.1786	0.2747	80.39%	64.98%
20	0.1356	0.2448	85.24%	69.52%
30	0.1294	0.2520	89.16%	67.41%

Table 2. Results for K-way shot learning

From the above table, we can see that training accuracies increase and training loss reduce consistently, especially for high values of K, rising from 80.39% to 89.16%. Such phenomenon can be evidenced from lots of literatures including some papers focusing on compare k way shot learning(Vinyals et al., 2016). However, the validation accuracies fail to appear the same trend, which arrive the highest when K is set to 20. Taking likely overfit into account, we chose 20 way in our third model.

3.3.2. Model Comparison

To visualize the task we worked on, The following two figures can represent clearly what classification question we try to solve. The basic CNN network used in baseline system got a single painting from the dataset like in Figure2, what we did in the last report is to classify paintings to their belonged classes like in Figure3, like clustering.

Figure 4. raw inputs

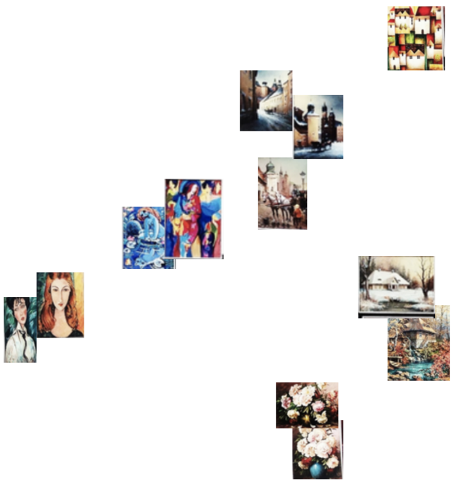
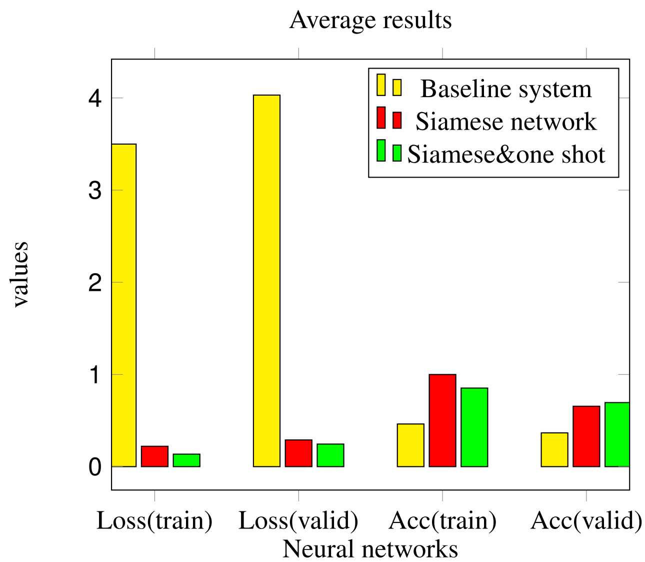


Figure 5. Classification

The difference between our baseline system and Siamese network is apparent, unsupervised learning in contrast to supervised learning. And it is easy to find in deep neural network work, supervised learning is very likely to perform better than unsupervised learning in most cases. For example, supervised learning consider more parameters, most importantly, the labels, which help to build reliable map(Zhao & Liu, 2007). However, the main comparison is focused on prediction accuracies of our proposed three models in deciding the authority of paintings. The below histogram graph illustrates their performance regarding traditional four criteria, training loss, validation loss, training accuracy and validation accuracy.

Interestingly, the importance of architecture can be seen here, Siamese networks even with 78720 pairwise examples during training process outperform the baseline system with 14.5773 million training pairwise samples in terms of all performance criteria listed in the histogram figure. Specially, The loss values in both training and validation set appear to decrease dramatically, with about 25 times and 15 times drops separately, not to mention that the training accuracy of the basic Siamese network is more than double our baseline system. Most importantly, there is an obvious rise with regards to the best average validation accuracy, increasing from nearly 36.64% in deep network to 64.98% in the second Siamese and 69.52% in the third model with the help of one shot learning.



J.Bromley (Bromley et al., 1994) in their work achieved 95% accuracy rate of validation dataset on Omniglot dataset that can be regarded as the transpose of MNIST using similar Siamese network together with one shot learning, which is really an excellent job. We suppose this gap is because we did not implement a wide range of techniques such as Bayesian hyperparameter optimization, layerwise learning rates/momentum and data augmentaion with distortions to enhance the return on validation accuracy like they did. Also, we trained less epochs just 200 and dataset.

In our preliminary experiments, we found the baseline is overfit apparently which may caused vary likely by the small subset we chose. Hence, we adopt a wide range of traditional strategies in order to reduce overfitting such as L2 Regularization, dropout and Batch Normalization. However, it is easy to see from the right two categories, it performs worse on tasks from the validition set than the train set. And this gap becomes wider in the basic siamese model. One possible explanation is that the training size of these two networks is way lower than baseline's painting pairs which result in even worse overfit again. It should be noted that the similar overfit is found in the basic Siamese (89% vs 65%) and complex Siamese network (85.24% vs 69.52%), but the later one's overfit behavior is alleviated to some extent, the work of Norouzi, M et al. which also apply one-shot method in their experiment has proved this opinion(Norouzi et al., 2013) .

We believe that the best model we chose ultimately failed to perform significantly better than our simple Siamese model since each of them is trained on the same dataset, and their weights converge to a similar optima. This perhaps suggest that one shot learning method here does not lead to a dramatic performance boost.

3.3.3. Test Performance

Based on the above experiments, the best model is the complex Siamese model with one shot learning through comparing validation accuracies. The held-out test set is used to test the best model's generation to unseen dataset. It is downloaded from this competition website, includes 23817 images in total. At the same time, this set consists of known authors and unseen authors in comparison with training data. We got around 57% test accuracy in this test set, which is more than 10% lower than validation accuracy and almost 30% less than training accuracy. Given this, the complex Siamese model does not generate well as we expected.

In order to find reasons for this, we analyzed the composition of the test set and found that the percentage of pairs belong to different authors is way bigger than the percentage of pairs belong to the same class, instead of the equal size of each class in training data. This may partly lead to the label 0 appears more often in predictions if our best model made a mistake because it is more likely to classify two samples to different authors' class.

4. Related work

As we stated in our introduction, the task of this project is to judge whether any two randomly selected paintings belong to the same artist. Fundamentally speaking, this problem is a problem to get the similarity between two images. According to a serials of papers of image processing in recent year, such as Krizhevsky et.al described in their work that CNN need less number of parameters and connection compared with standard feed-forward neural networks, so it is much easier to train, and their theoretically-best performance is likely to be only slightly worse(Krizhevsky et al., 2012). Besides, as Bottou et.al discussed before, Convolutional neural networks are a good example of an idea inspired by biology that resulted in competitive engineering solutions that compare favorably with other methods(LeCun et al., 1995). According to the experiment experience and the knowledge we got last semester, we apply such popular image processing deep neural network CNN network in our interim project model.

Siamese network is proposed by J.Bromley et.al(Bromley et al., 1994) in 1994. Actually J.Bromley proposed this method aiming to tackle the signature verification problem by measuring the similarity between two signatures. It solves the nearly same problem in our project which is to measuring the similarity between two objectives. In this situation, our model probably suits the Siamese CNN neural network structure well. Also, such combination is very common in both previous and current works. Koch G et.al have combined siamese network and convolutional network to finish image recognization(Koch et al., 2015). They established a siamese convolutional neural network to train Omniglot database. They gave the model two images once and train it to predict whether the two images belong to the same class, and then when doing a one-shot classification task, the network compares the test image with every image in the support set to pick up one image that is most likely to belong to the same class.
The research of C. R. Johnson et.al.(Johnson et al., 2008) summarized the experiment result of several groups based on the same dataset using 16-b gray-scale digitized representations of very high spatial resolution (196.3 dpi) of 101 paintings, mostly by van Gogh. In their research they have analyzed three typical different wavelet settings to construct features for supervised learning.

The first project applied one-shot method in image processing work is the work finished by Fei.L et al. Fei.L et al who have developed a framework using one-shot method to realize image classification. Their work proved that one-shot method can really increase performance on forecasting future ones when very few examples are available from a given class (Fe-Fei et al., 2003).
We got the idea of applying One-shot method mainly from the paper of Koch.G et.al(Koch et al., 2015). There is a common sense that if we train the network with more training examples, the more features the network will learn. But, the overfit may be likely to be caused in such situations. According to the paper of Koch.G et.al, They have trained their model with a few examples exist for some classes using the combination of Siamese network architecture and one-shot learning method, their final result verified that applying one-shot method could indeed increase the performance of model with few training examples.
There are also some other similar researches have been done in recent year.Bart et.al learn novel classes from a single example applying one-shot methodology(Bart & Ullman, 2005). Vinyals et.al (Vinyals et al., 2016) have established a matching networks for one-shot learning.

5. Conclusion

Although there are many literatures in this regard like handwriting verification, very little work has been done in painting authority identification. In this paper, we explored different models like Siamese convolutional neural netowk architectures to verify authorship of paintings. We first adjust CNN parameters aiming to find the best settings of the model in prediction accuracy. Then we confirmed the values for discouraging overfit in various approaches such as dropout, L2 refularization and Batch normalization. This two study directions have been examined in a large amount experiments with reference to our last work. And we adopt these optimal values in our latter three models.

Next, we compare different architectures to decide which one provided the most inspiring validation accuracy based on the same dataset and CNN settings. Beyond our baseline model, we explore a simple Siamese network and combine this model with one shot learning method which is the best performing model. It was trained on pairs coming from the same dataset but such pairwise data is way less than that in the baseline system. Despite this, it was able to achieve nearly 70% validation accuracy, almost three times than that of baseline model. However, its return on test accuracy did not very conform to our expectation, only 57%. The main cause for this we suspect may be the different weights in painting pairs, equal percentage in training set but larger data percentage in different class than the same painter class. Thus, we suppose that this model can be further improved regarding enhancing its generation ability to unbalanced distribution between two classes rather than limiting in equal pair numbers in each one.

In conclusion, we have shown that the Siamese convolutional neural network combining one shot learning can not only perform well using a small data set but also discourage overfit behavior. It gives an encouraging exploring way of classification problem in practical identification of painting authority and the model we conducted is able to be developed better, using a wide rage of tricks in boosting its learning ability. For example, Bayesian hyper-parameter optimization was taken into account when we carried out experiments. However, we failed to apply it due to limited time. Potential improvements of this model could become an interesting extension of our project.

References

Anderson, Robert C. Paintings as an investment. *Economic Inquiry*, 12(1):13–26, 1974.

Bart, Evgeniy and Ullman, Shimon. Cross-generalization: Learning novel classes from a single example by feature replacement. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pp. 672–679. IEEE, 2005.

Bishop, Christopher. *Machine Learning and Pattern Recognition*. Springer-Verlag New York, 2014.

Bromley, Jane, Guyon, Isabelle, LeCun, Yann, Säckinger, Eduard, and Shah, Roopak. Signature verification using a" siamese" time delay neural network. In *Advances in Neural Information Processing Systems*, pp. 737–744, 1994.

Chopra, Sumit, Hadsell, Raia, and LeCun, Yann. Learning a similarity metric discriminatively, with application to face verification. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pp. 539–546. IEEE, 2005.

Fe-Fei, Li et al. A bayesian approach to unsupervised one-shot learning of object categories. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1134–1141. IEEE, 2003.

Hadsell, Raia, Chopra, Sumit, and LeCun, Yann. Dimensionality reduction by learning an invariant mapping. In *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, volume 2, pp. 1735–1742. IEEE, 2006.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

Johnson, C Richard, Hendriks, Ella, Berezhnoy, Igor J, Brevdo, Eugene, Hughes, Shannon M, Daubechies, Ingrid, Li, Jia, Postma, Eric, and Wang, James Z. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4), 2008.

Koch, Gregory, Zemel, Richard, and Salakhutdinov, Ruslan. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

LeCun, Yann, Bengio, Yoshua, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

Norouzi, Mohammad, Fleet, David J, and Salakhutdinov, Ruslan R. Hamming distance metric learning. In *Advances in neural information processing systems*, pp. 1061–1069, 2012.

Norouzi, Mohammad, Mikolov, Tomas, Bengio, Samy, Singer, Yoram, Shlens, Jonathon, Frome, Andrea, Corrado, Greg S, and Dean, Jeffrey. Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*, 2013.

Vincent Dumoulin, Francesco Visin. A guide to convolution arithmetic for deep learning. 2016.

Vinyals, Oriol, Blundell, Charles, Lillicrap, Tim, Wierstra, Daan, et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pp. 3630–3638, 2016.

Zhao, Zheng and Liu, Huan. Spectral feature selection for supervised and unsupervised learning. In *Proceedings of the 24th international conference on Machine learning*, pp. 1151–1157. ACM, 2007.