Labeling Images with Thematic and Emotional Content

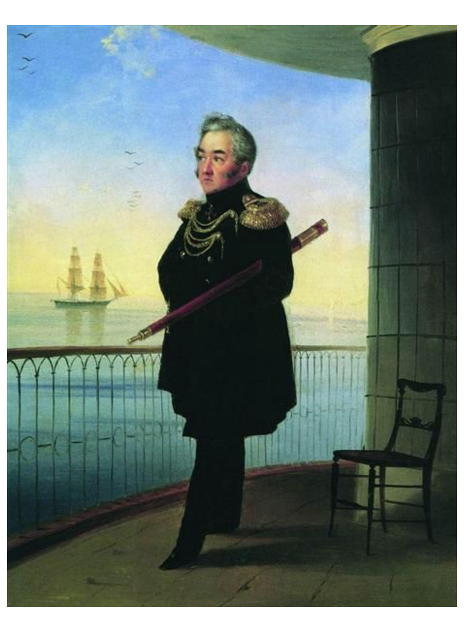
Bardia Beigi {bardia}, Soroosh Hemmati {shemmati}
Department of Computer Science, Stanford University

Introduction

This project is motivated by the authors' deep appreciation of visual arts and their attempt to find a way to use deep learning to classify paintings. The outcome of this project could be used to differentiate and identify features that make up how experts classify paintings without any prior knowledge on part of the user. The filters produced in the model can be used to gain a deeper understanding of what can be used to decide a painting's style and genre. Models could also be developed to identify similarities and differences between stroke patterns, strengths, frequency, and overall identification of different painters' works.

Problem Statement

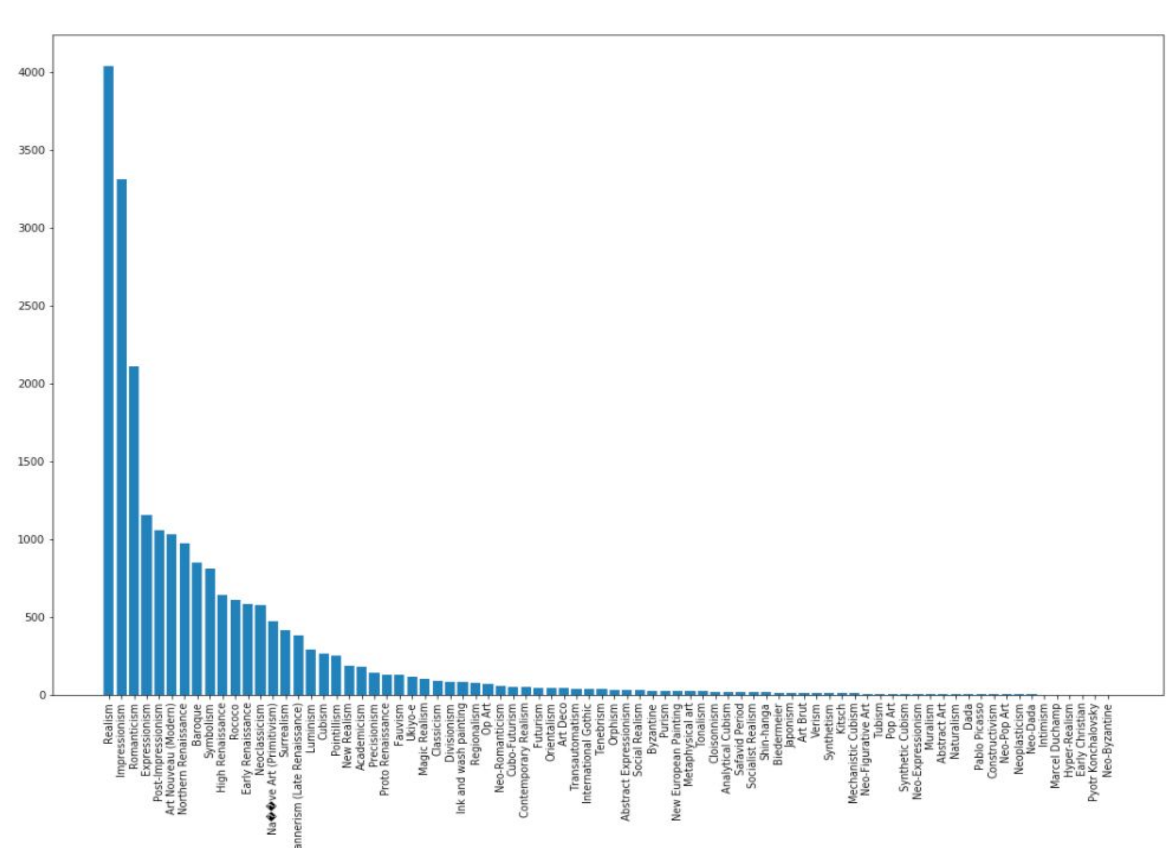
The goal of this project is to classify a set of over 35000 paintings based on style and theme. Style consists of labels such as "Romanticism", "Baroque", etc. and theme includes classes such as "portraits", "animals", etc. The images all have other data such as genre, artist, theme, and style. We will explore the effects of various CNN designs on this dataset. Due to images having clear labels, evaluation is done based on what percentage of images get the correct classification.



Style:	Romanticism	Neoplasticism
Theme:	male-portrait	animals

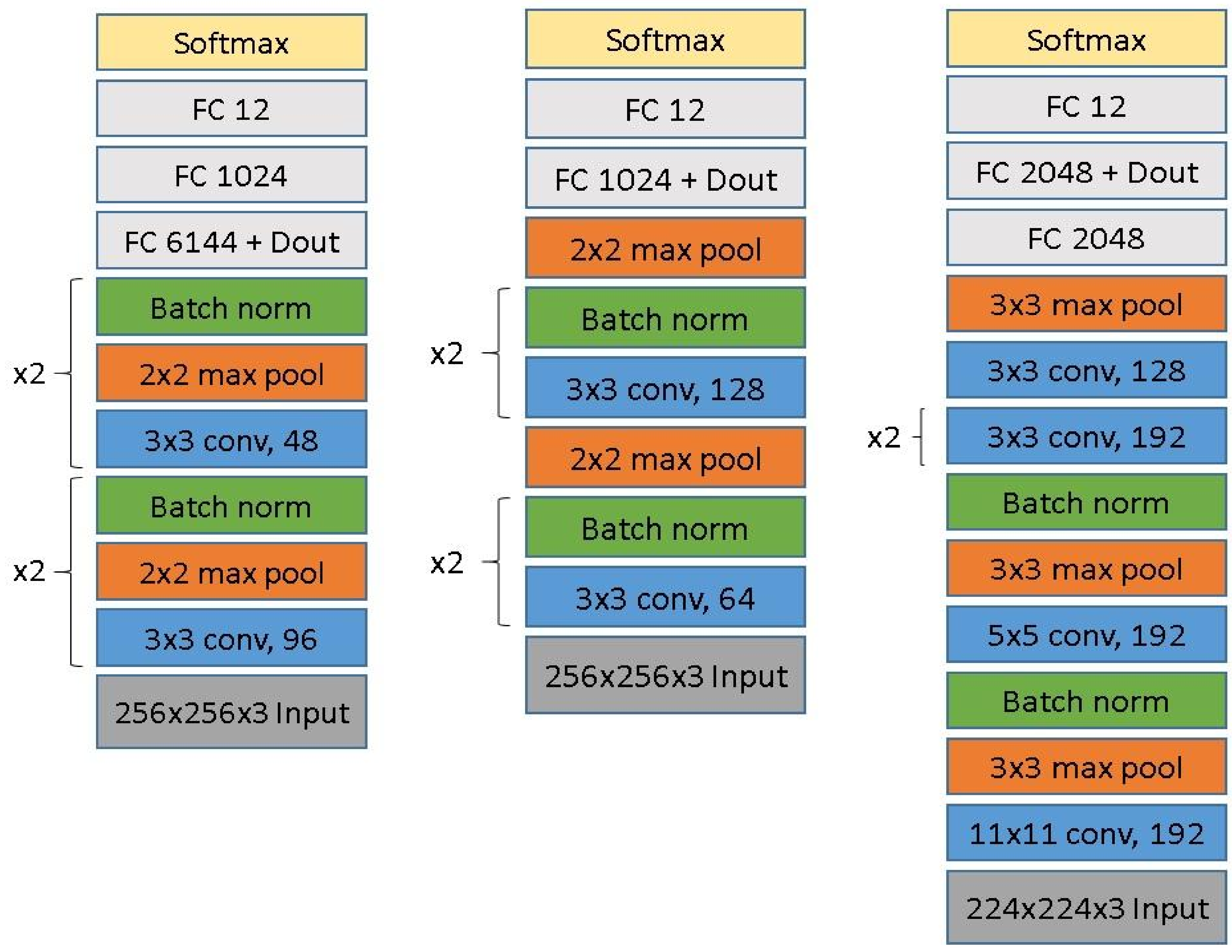
Data

As previously mentioned, the dataset contains over 35000 images taken from all images have labels such as style and theme, taken from WikiArt. The distribution of classes for each label is highly variant, however. For instance, stylistic labels have a highly non-uniform shape as shown below. Same is true for thematic labels even though the distribution is slightly better. Preprocessing was done to turn the csv file containing links to useful images for training and classification.



Models and Algorithms

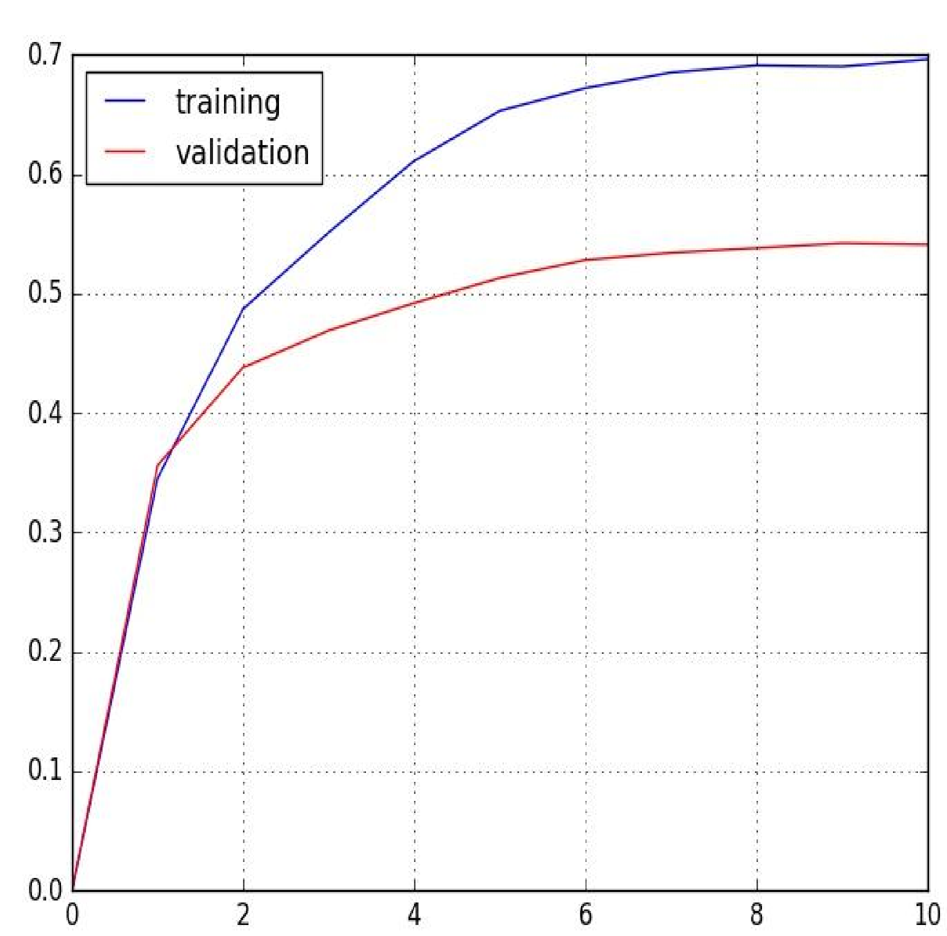
- Customized Design:
- VGGNet:
- AlexNet:



Due to the high amount of variation in the number of classes and to equalize the contribution of each class to the training pool, classes with fewer elements were reflected and subsampled multiple times. The data was also shuffled to ensure works of painters would spread out across the whole dataset. We also only worked with the top 12 most represented classes.

Results

Although investigations are ongoing, the best result gathered so far came from the customized design in which we got almost **70% train, and 54% validation accuracy** for style classification on the 35000 images in the dataset.



	Training Accuracy (%)	Validation Accuracy (%)
Customized Design	**69.1**	**54.2**
VGGNet	**63.2**	**48.4**
AlexNet	**61.8**	**44.8**

Conclusion & Future Directions

As evidenced in the results, there is clearly room for improvement. Going forward, we will be exploring deeper and heavier models capable of learning more, ensembles of models, and tricks in which a picture is subsampled, each subsample classified, and the image is classified using the most represented class. Image classification will also be explored based on other labels such as theme. If time permits, we will also explore nearest neighbor classifications which might prove useful due to the nature of the problem. Due to the high amount of diversity in the images, this problem is more complex than image classification using CIFAR-10. These models will be explored in more detail and presented in the final report.