Comparison of Machine Learning Techniques for Artist Identification

Jennie Chen, Andrew Deng
{jchen437, andrewde} @ stanford.edu

Objective

Many paintings have unknown or highly contested artists, and experts in the field need a long time to learn various styles before answering these questions. Machine learning can be used to reduce time and effort. In this project, we will be comparing two techniques in terms of performance and ease of use for the task of artist classification.

Dataset

- Obtained from Kaggle competition "Painters by Numbers"
- Narrowed down to 15 artists with at least 450 paintings each for a total of 7462 paintings
- Color images, resized to 256 by 256 pixels
- Divided into training/validation/testing in a 80/10/10 split
- Data augmentation (rotation, zooming, etc.) for CNN model

Models

SVM

- Multi-class SVM with RBF kernel
- PCA with 100 components to condense feature space
- Trained models across various combinations of features

CNN

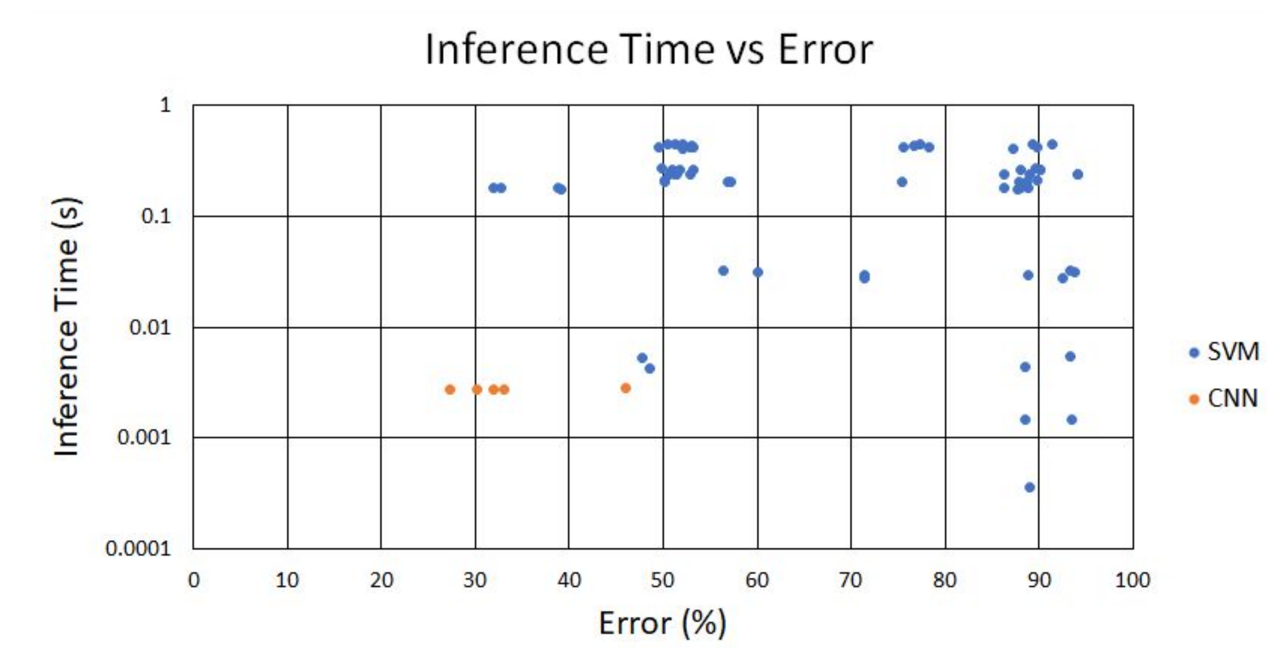
- Architecture based on winning submission to the official Kaggle competition [1]
- 11 convolutional layers
- 2 dense layers
- Batch normalization after each computation layer
- Dropout layers after last max-pool and before softmax layer with keep probability 0.5
- ReLU activation functions
- Adam optimizer with cross-entropy loss

Input Image	(256, 256, 3)
Convolution (16, 3, 3)	(256, 256, 16)
Convolution (16, 3, 3)	(256, 256, 16)
Max-Pool (2, 2)	(128, 128, 16)
Convolution (16, 3, 3)	(128, 128, 16)
Max-Pool (2, 2)	(64, 64, 16)
Convolution (32, 3, 3)	(64, 64, 32)
Max-Pool (2, 2)	(32, 32, 32)
Convolution (128, 3, 3)	(32, 32, 128)
Max-Pool (2, 2)	(16, 16, 128)
Convolution (256, 3, 3)	(16, 16, 256)
Max-Pool (2, 2)	(8, 8, 256)
Flatten	(16384,)
Dense	(2048,)
Dense	(2048,)
Softmax	(15,)

Results

Model	Training Accuracy (%)	Test Accuracy (%)	Train Time (s)	Inference Time (ms)	Features Used
LR*	67.4	61.2	1.86	179.8	GIST, Color Hist
SVM	**97.9**	**68.1**	**6.78**	**180.2**	**GIST, Hu Moments**
SVM	97.8	67.3	6.69	181.0	GIST, Color Hist
SVM	97.1	61.2	5.12	178.4	GIST, Color Hist, Hu Moments
CNN	**81.3**	**74.7**	**28921**	**12.2**	**NA**

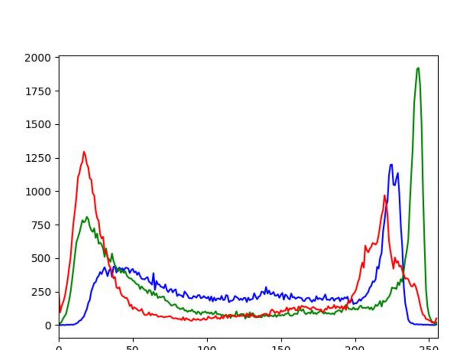
* Baseline model - logistic regression with best combination of features



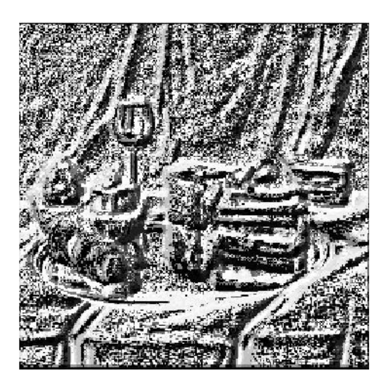
Feature Extraction

Most useful features:

- GIST: features to estimate the "shape of a scene" in terms of five different criteria (e.g. naturalness, openness)
- Hu Moments: descriptors of object shapes in image
- Color Histogram: distribution of colors throughout image



Color Histogram

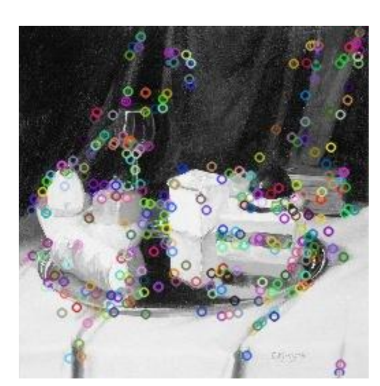


Local Binary Pattern

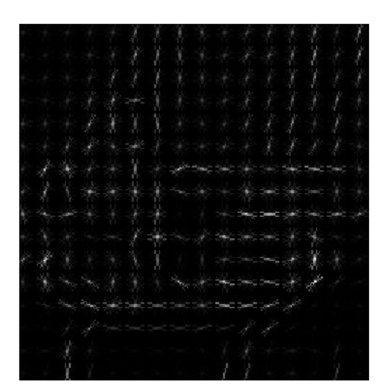
Other features extracted:

- Scale-Invariant Feature Transform (SIFT)
- Histogram of Oriented Gradients (HOG)
- Local Binary Pattern (LBP)
- ORB
- Haralick Texture

Original Image



SIFT Keypoints



HOG

Future Work

- Further attempts at addressing SVM overfitting
  - Already tuned regularization parameter C
  - Explore options such as early stopping
- Explore Classemes and Picodes as additional features
- Try transfer learning from existing models like AlexNet or ResNet

Discussion

- Training CNN took significantly longer (7-8 hours).
- Feature extraction took ~44 minutes for the training data.
- Inference times on same CPU were lower on the CNN for better/similar accuracy.
- Using a SVM requires a lot of overhead in tuning feature combinations compared to feeding images into the CNN.
- Only a subset of features (GIST, Hu, Color Hist) proved useful; some (Haralick) had detrimental effect on accuracy

References

[1] GitHub. 2018. Winning solution for the Painter by Numbers competition on Kaggle - inejc/painters. Retrieved from http://github.com/inejc/painters.