Automatic Classification of Woodcuts and Copperplate Engravings

Michael Cormier[1], Sarah Park[1] & Lauren Beck[2]

[1]Dept. of Mathematics and Computer Science, Mount Allison University; [2]Dept. of Visual and Material Culture Studies, Mount Allison University

INTRODUCTION & MOTIVATION

- Woodcuts and copperplate engravings are two methods used to create illustrations
- Distinguishing these methods require manual labelling by experts
- Reducing manual labelling efforts will allow these experts to allocate their time to ambiguous cases and other tasks

Gradient magnitudes were expected to reveal the differences in how these engravings are printed, as the methods used create differences in lines printed.

APPROACH

We present two Bayesian statistical machine learning approaches for distinguishing woodcut from copperplate engravings. Manual classification efforts are resource-demanding and off-loading these to algorithms allow experts to work on cases requiring efforts (e.g. ambiguous cases)

A *Bayesian* approach was chosen over a CNN for several reasons:

- Interpretability: Probabilistic estimates are preferred for industry experts
- Flexibility: We can expand the model to account for other features
- Sample size: The dataset used was not sufficient for a CNN approach
- *A priori* reasons to suspect the gradient will be sufficient: Physical processes were expected to produce gradient distinctions
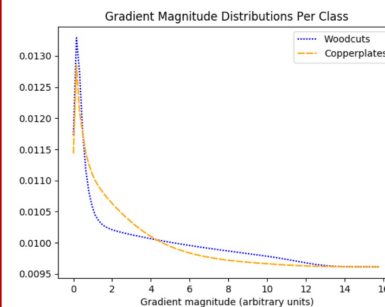
DATASET

Woodcuts: characterized by fewer, coarser lines

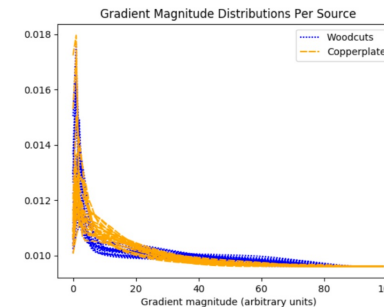*Copperplate engravings*: characterized by many fine lines

Pre-Processing
- 400 images from 40 sources
- All images were were scaled to 600 PPI

Per Class:
- Distinct distributions for each class

Per Source:
- Variation within a source

ALGORITHMS

Low-level:
- Learn a distribution of gradient magnitudes for each class
- Classify image using the probability that the gradient was of that class

High-level:
- Compute the mean, age, variance, skew, and kurtosis of the distribution of gradient magnitudes
- Learn a non-parametric joint distribution for each class

RESULTS

Accuracy
- Both algorithms show a high accuracy for performance
  - Maximum 88.8% on the validation set

Hyperparameters:
- Accuracy remains stable
  - *Low-level:* Number of bins in histogram representation
  - *High-level:* Kernel width

CONCLUSIONS

- Both algorithms are effective and robust and consequently suitable for distinguishing woodcuts from copperplate engravings
- Probability estimates are easily interpretable by experts and this structure allows other features (e.g. publication date)