

---

# Classification of Engraved Illustrations using a Statistical Machine Learning Approach

---



Sarah Park  
Supervisor: Dr. Michael Cormier  
Mount Allison University  
Department of Mathematics and Computer  
Science

Science  Atlantic  
Computer Science  
Mathematics & Statistics | Conference

October 26, 2019  
Dalhousie University, Halifax, NS

# PROBLEM & MOTIVATION

- Classification of engraved illustrations

# PROBLEM & MOTIVATION

- Classification of engraved illustrations
- Creation of a useful database with metadata
  - “Search” function

# PROBLEM & MOTIVATION

- Classification of engraved illustrations
- Creation of a useful database with metadata
  - “Search” function
- Use a computer vision approach to increase efficacy and reduce human labelling



## BACKGROUND: THEORY

## BACKGROUND: THEORY

- Copperplate engravings
  - Many fine lines

## BACKGROUND: THEORY

- Copperplate engravings
  - Many fine lines



## BACKGROUND: THEORY

- Copperplate engravings
  - Many fine lines
- Woodcut engravings
  - Fewer coarse lines





## BACKGROUND: THEORY

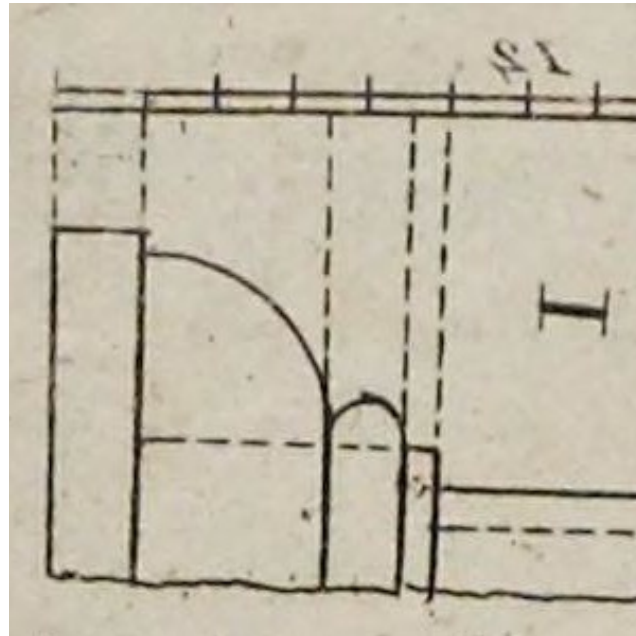
- Copperplate engravings
  - Many fine lines
- Woodcut engravings
  - Fewer coarse lines



# BACKGROUND: THEORY

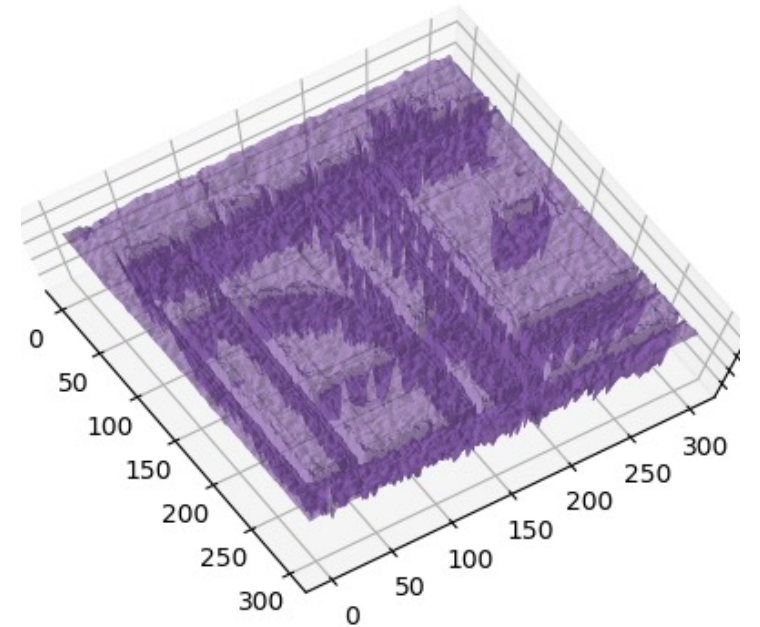
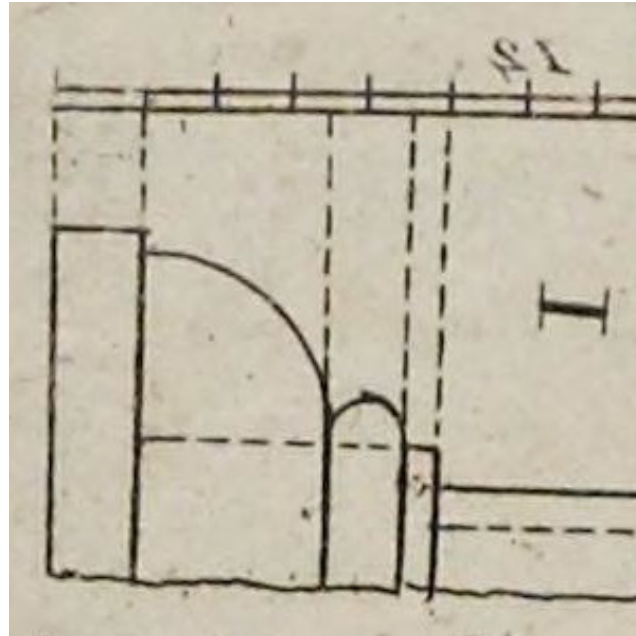
# BACKGROUND: THEORY

- Images formed by lines
  - dark edges and light backgrounds



# BACKGROUND: THEORY

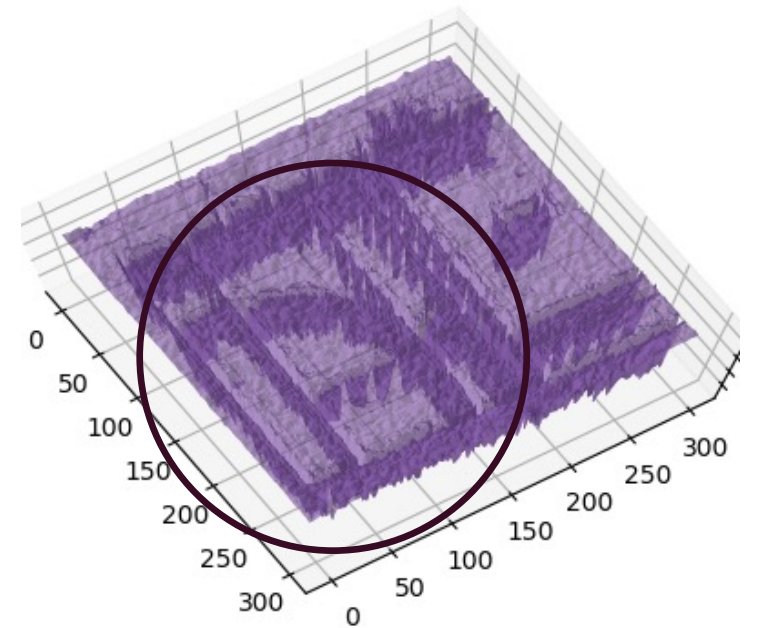
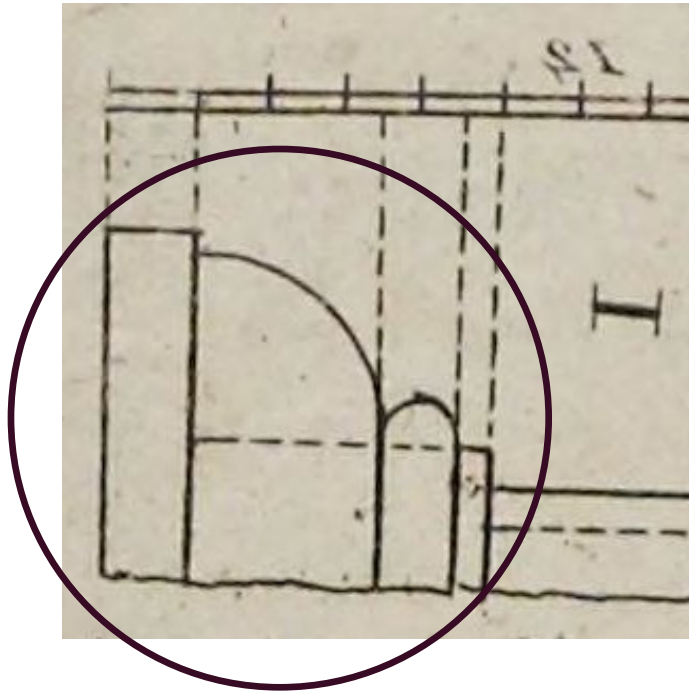
- Images formed by lines
  - dark edges and light backgrounds
- This contrast can be represented via the gradient





# BACKGROUND: THEORY

- Images formed by lines
  - dark edges and light backgrounds
- This contrast can be represented via the gradient





# BACKGROUND: APPROACH



## BACKGROUND: APPROACH

- Distinguishing low-level features

## BACKGROUND: APPROACH

- Distinguishing low-level features
- Bayesian approach
  - Results can be easily interpreted
    - E.g. 70% chance the illustration is a woodcut



## BACKGROUND: APPROACH

- Distinguishing low-level features
- Bayesian approach
  - Results can be easily interpreted
    - E.g. 70% chance the illustration is a woodcut
  - Allows for other properties to be added

## BACKGROUND: APPROACH

- Distinguishing low-level features
- Bayesian approach
  - Results can be easily interpreted
    - E.g. 70% chance the illustration is a woodcut
  - Allows for other properties to be added
  - Threshold of uncertainty such that images can be flagged for manual review

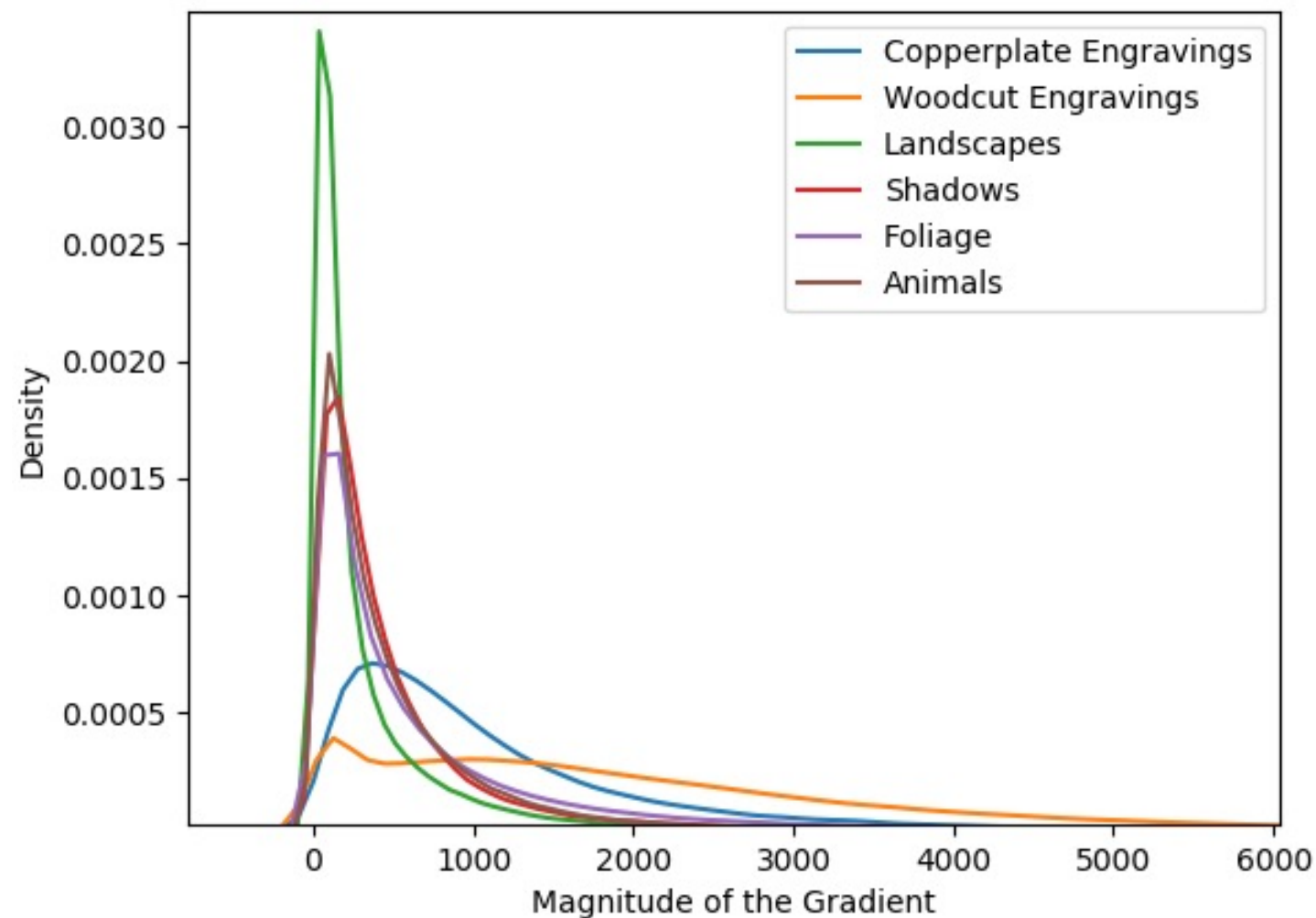
# COMPUTER VISION

# COMPUTER VISION

- Line properties, edge detection and gradient descent
- Generate histogram of the gradient magnitudes for each image

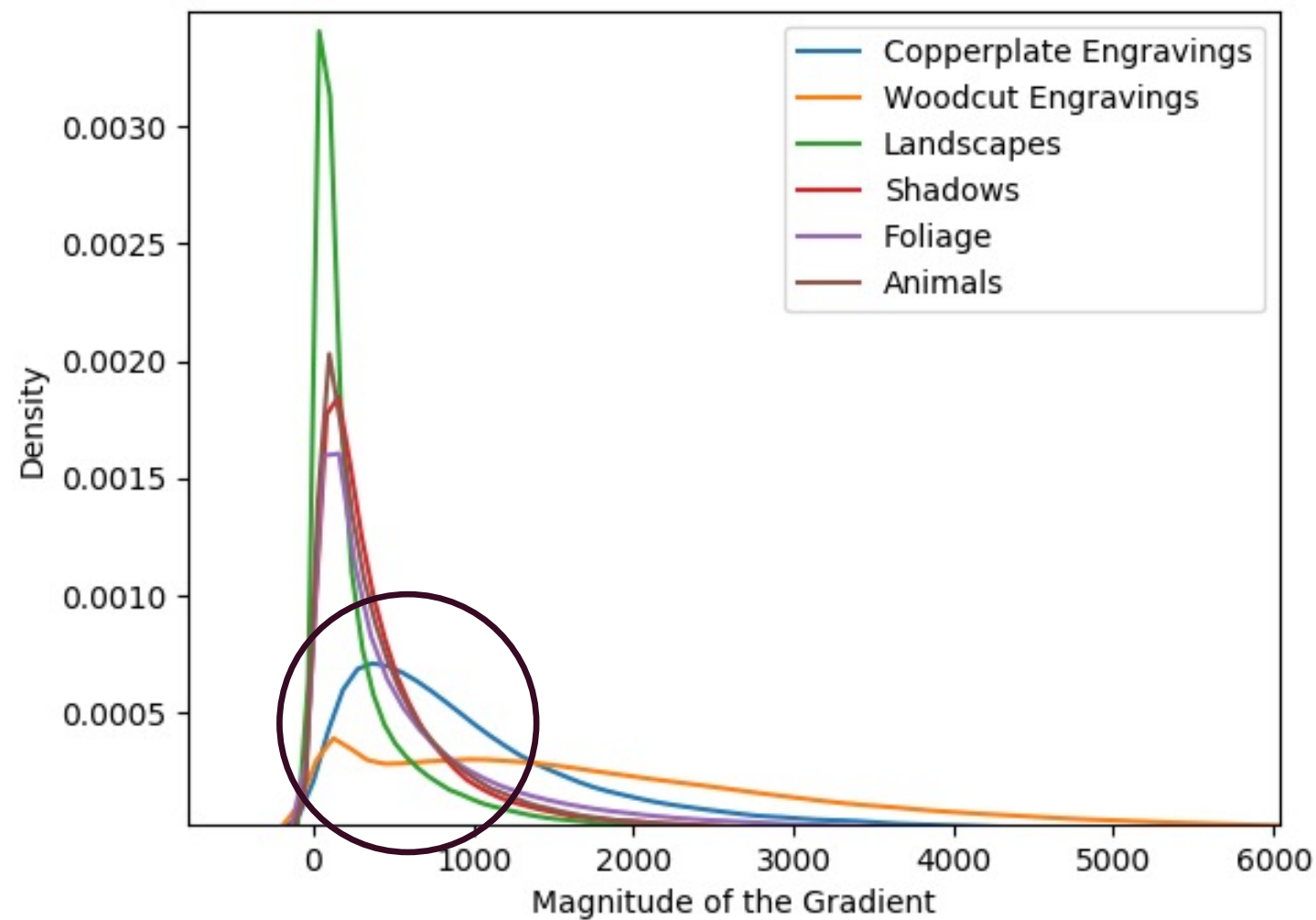
# COMPUTER VISION

- Line properties, edge detection and gradient descent
- Generate histogram of the gradient magnitudes for each image



# COMPUTER VISION

- Line properties, edge detection and gradient descent
- Generate histogram of the gradient magnitudes for each image



# METHOD

# METHOD

- Preprocessing
  - Images read in as grayscale
  - Resized
  - Scaling
  - Flattened to create histogram & assigned a label



# METHOD

## METHOD

### Learning

- Learn distribution of gradient magnitudes for each class

# METHOD

## Learning

- Learn distribution of gradient magnitudes for each class

## Classification

- Take histogram of gradient magnitudes
- Calculate probability of generating it from each class
- Bayes' theorem

# METHOD

$H = h$ : histogram of the gradient magnitudes is  $h$

$C = c$ : image class is  $c$

$$P(C = c | H = h) = \frac{P(H = h | C = c) P(C = c)}{P(H = h)}$$



# PRELIMINARY RESULTS



# PRELIMINARY RESULTS

- McGill test data
  - (n = 201) 79% - (n = 429) 84% accuracy

# PRELIMINARY RESULTS

- McGill test data
  - (n = 201) 79% - (n = 429) 84% accuracy
- Experimental data (Woodcut engravings vs Copperplate engravings)
  - (n = 245) 95% accuracy

# LIMITATIONS



# LIMITATIONS

- Scanning artifact and texture confounds

# LIMITATIONS

- Scanning artifact and texture confounds
  - Converting to grayscale
    - Strong texture tend to stretch the low-gradient peak
    - Light texture tends to condense the low-gradient peak

# LIMITATIONS

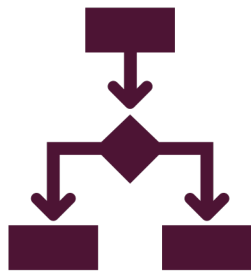
- Scanning artifact and texture confounds
  - Converting to grayscale
    - Strong texture tend to stretch the low-gradient peak
    - Light texture tends to condense the low-gradient peak
- Sample size and potential for overfitting

# LIMITATIONS

- Scanning artifact and texture confounds
  - Converting to grayscale
    - Strong texture tend to stretch the low-gradient peak
    - Light texture tends to condense the low-gradient peak
- Sample size and potential for overfitting
  - Pixel scaling

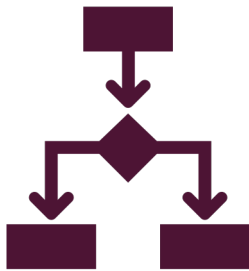
# CONCLUSIONS

# CONCLUSIONS



Classify images based on  
low-level image properties

# CONCLUSIONS

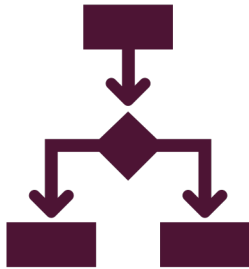


Classify images based on  
low-level image properties



Preliminary results are  
promising

# CONCLUSIONS



Classify images based on low-level image properties



Preliminary results are promising



Further work to increase the validity of results