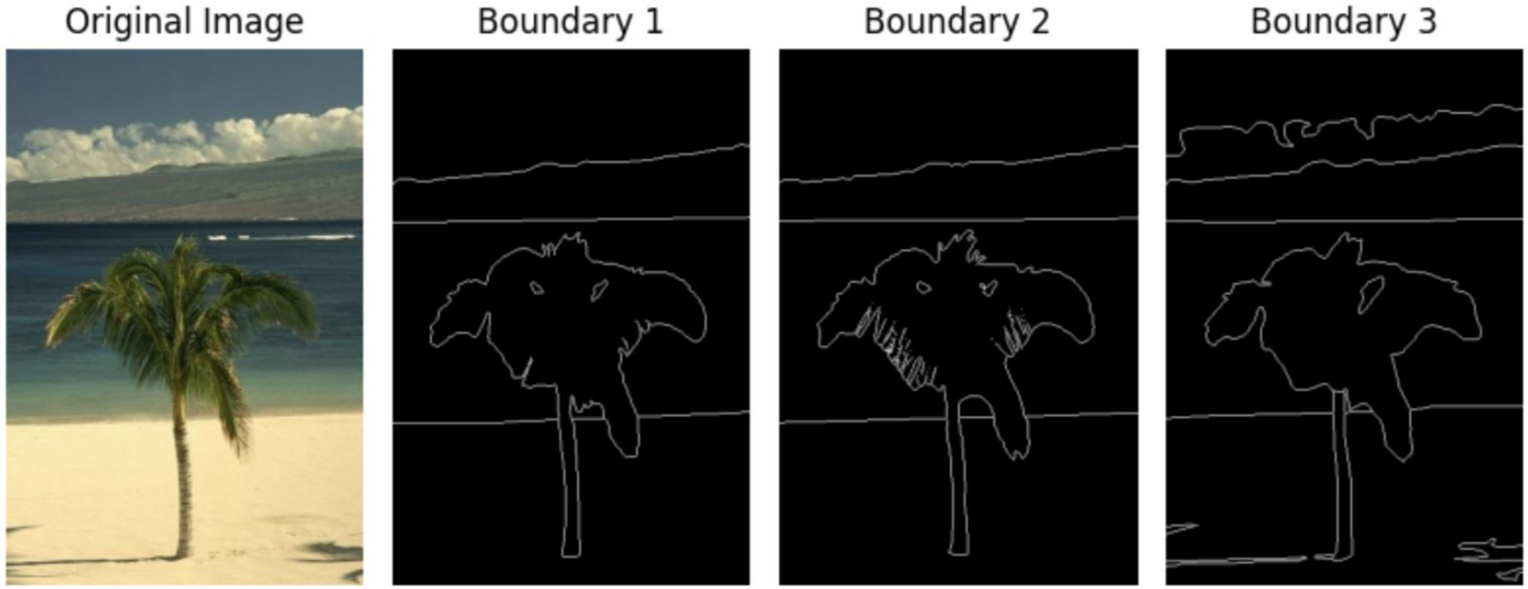
Boundary Detection

Group 5: Qin Tong Bettina Tee, Daso Jung, Yan Renyu

Contents

Problem Statement

Predict boundary score maps for given image, where each pixel's value (between 0.0 and 1.0) indicates the likelihood of being part of an object boundary.



Approach taken to solve the problem

1. Feature + Target Extraction
2. Feature Importance
3. Parameter Prediction
4. Image Processing
5. Edge Detection

1. Feature + Target Extraction

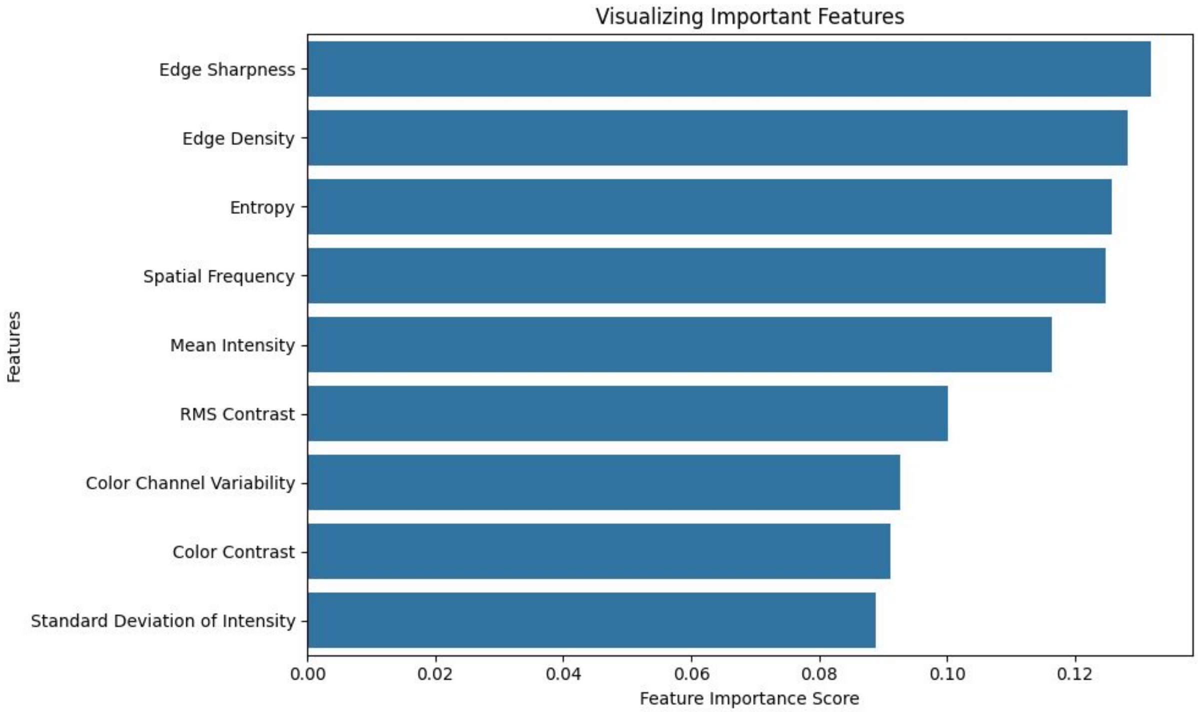
Features extracted:

1. Edge Density
2. Edge Sharpness
3. Colour Contrast:
4. Root Mean Square Contrast
5. Entropy
6. Mean Intensity
7. Standard Deviation of Intensity
8. Colour Channel Variability
9. Spatial Frequency

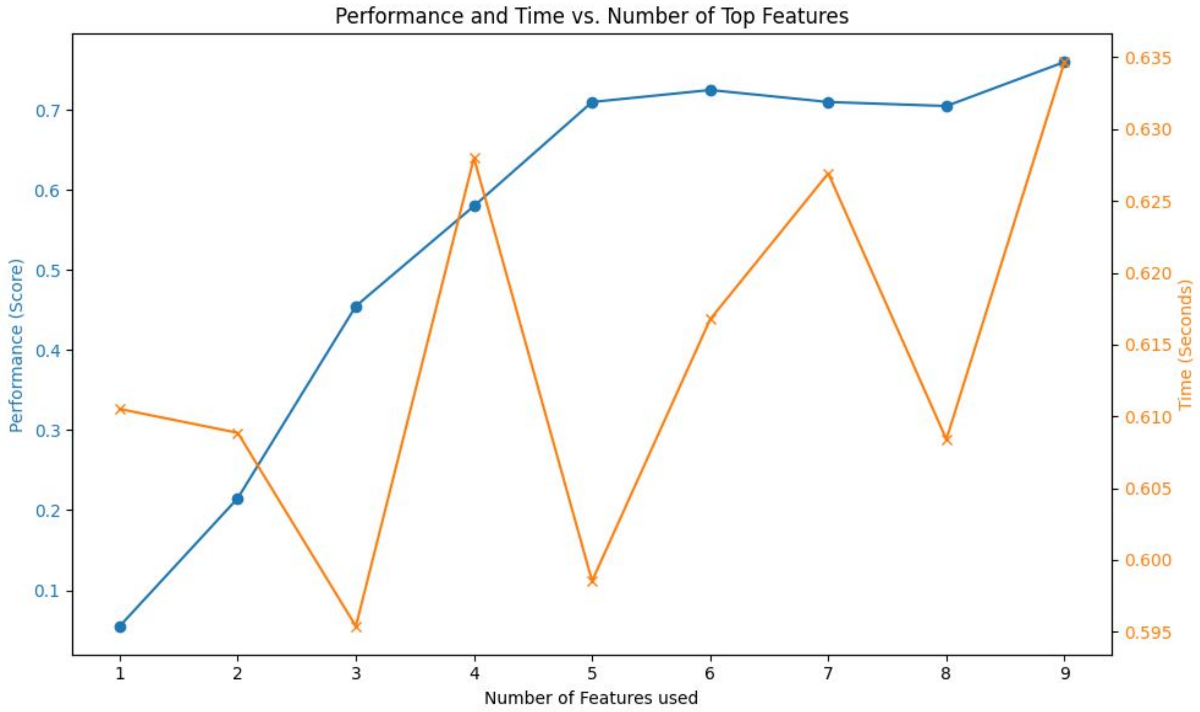
Target (parameters for preprocessing):

1. Gaussian Blurring:
   a. blur_amount
   b. sigma
2. Contrast Enhancing:
   a. contrast_factor

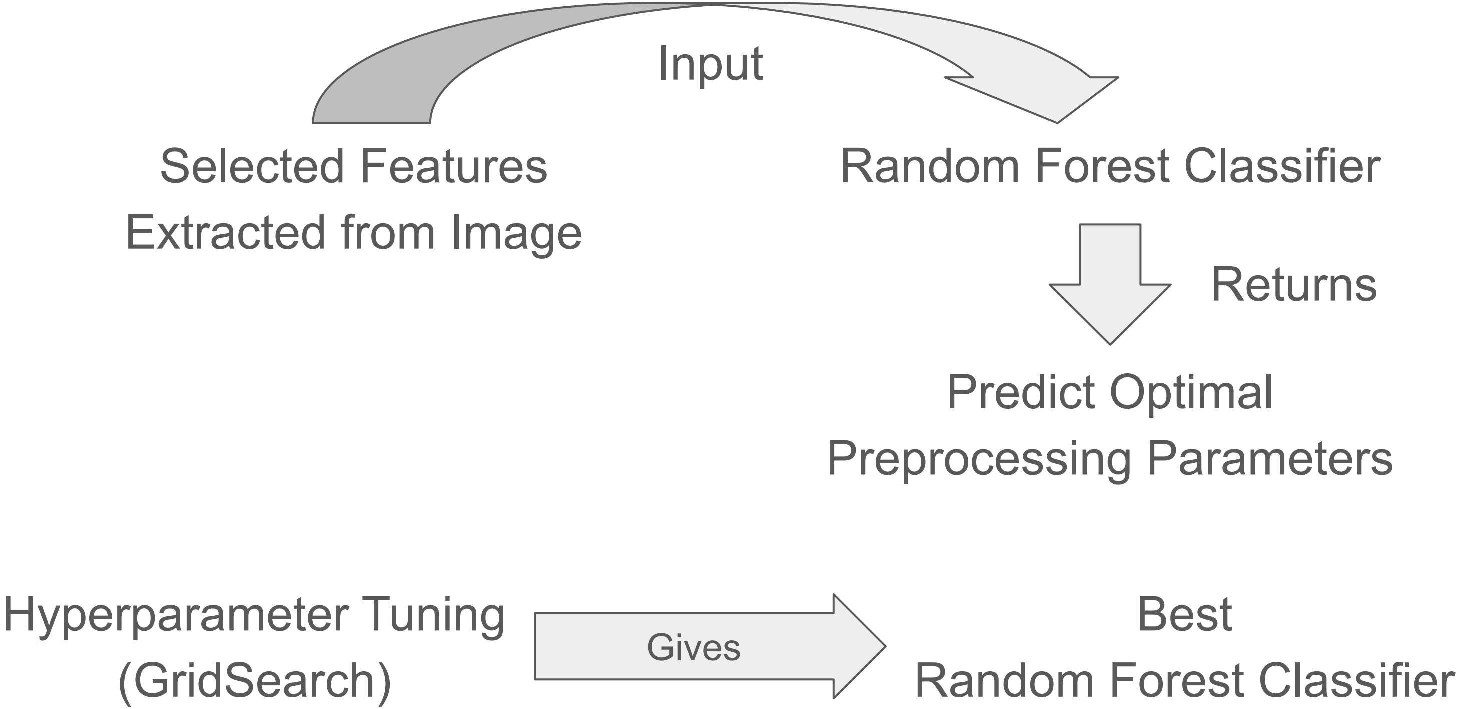
2. Feature Importance



2. Feature Importance



3. Parameter Prediction



4. Image Preprocessing

→

Gaussian Blurring

+

Contrast Enhancing

Optimized Parameter

1. Sigma
2. Blur amount

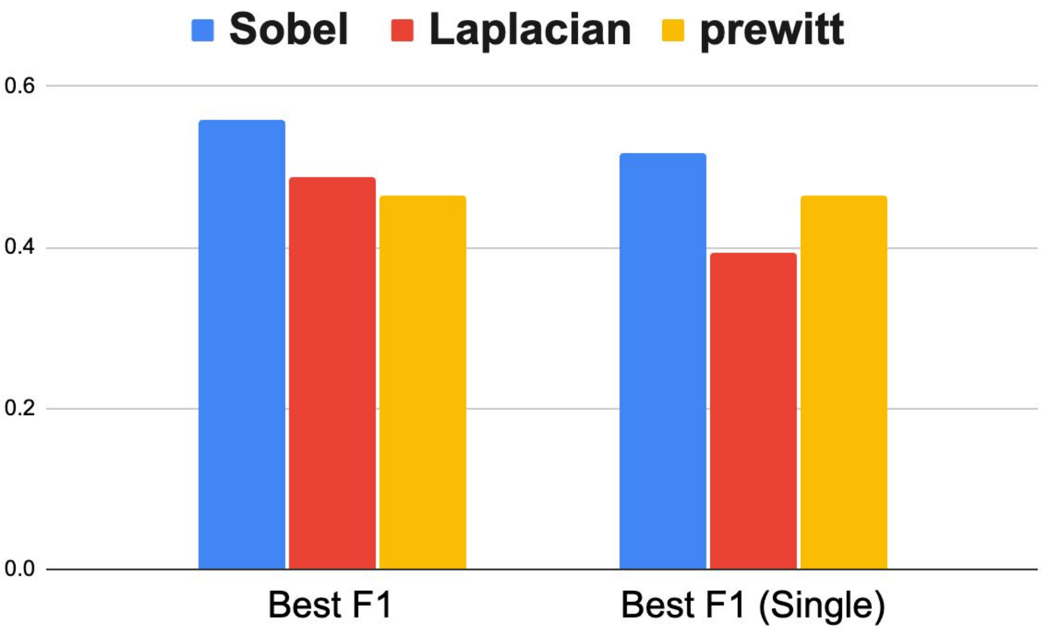
Optimized Parameter

1. Contrast Factor

5. Edge Detection

Boundary 1

Sobel Operator



Key Findings and Observations

Key metrics

- F1-score: Harmonic mean of precision and recall
- Best F1: Average F1-score using optimal threshold for each image
- Best F1 (Single): Average F1-score using a single optimal threshold for all images
- Recall: proportion of actual edges correctly identified
- Precision: Proportion of correctly predicted edges among all classified edges

Impact of Preprocessing

	Average F1-score	Best F1	Best F1 (Single)
Without preprocessing	0.54	0.56	0.52
With preprocessing	0.64	0.65	0.63

Improvement:

- 18.5% increase in average F1-score (from 0.54 to 0.64)
- Significant impact in Sobel operator performance.

Training Set Results

- Metrics:
  - Average F1-score: 0.57
  - Best F1: 0.59, Best F1 (Single): 0.56
  - Recall: 0.67, Precision: 0.52
- Range Across Images:
  - Recall: 0.4 to 0.95
  - Precision: 0.2 to 0.9
- Insights
  - Optimized thresholds marginally improve F1-scores
  - High recall indicates most edges were detected
  - Lower precision suggests more false positives

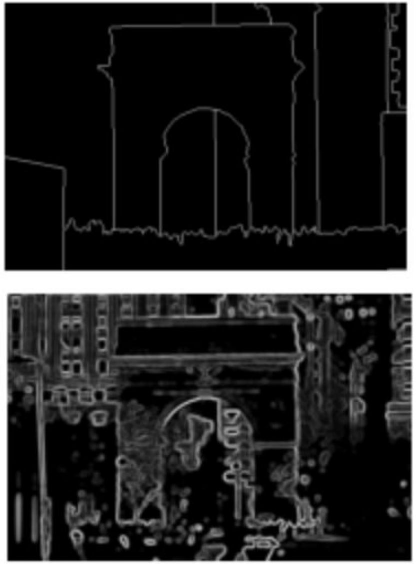
Test Set Results

- Metrics:
  - Average F1-score: 0.54
  - Best F1: 0.55, Best F1 (Single): 0.52
- Insights
  - Image-specific thresholds improve results
  - Slight performance drop from training to test set
  - Training F1-score: 0.57, Test F1-score: 0.54
  - Gap of 0.03 suggests minor overfitting

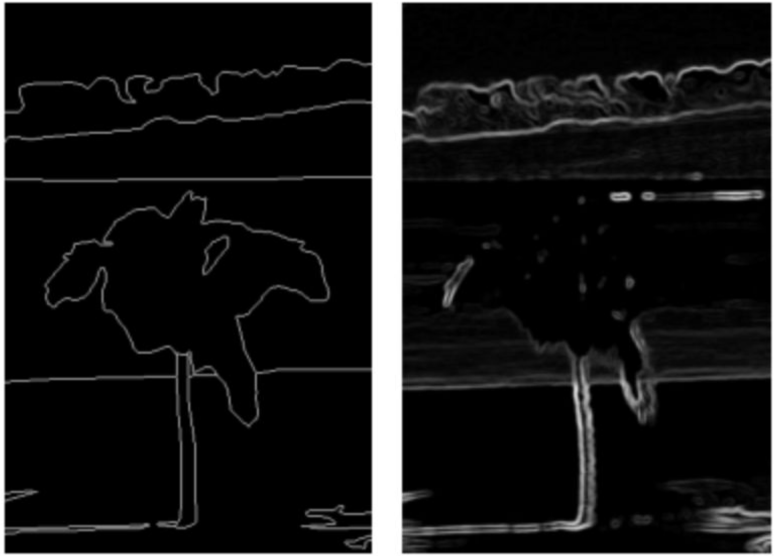
Live Demo

Challenges

1. Different images characteristics



Too much detail



Overly simplified

Challenges

2. Very wide grid search space, a long time to run grid search → Bayesian Optimisation (less time + computation intensive)

```
# search space of paramters of blur_amounts, sigmas, contrast_factors
search_space = [
    Categorical([3, 5, 7, 9]), # blur_amount for red
    Categorical([3, 5, 7, 9]), # blur_amount for green
    Categorical([3, 5, 7, 9]), # blur_amount for blue
    Categorical([1.0, 3.0, 5.0, 7.0, 9.0]), # sigma for red
    Categorical([1.0, 3.0, 5.0, 7.0, 9.0]), # sigma for green
    Categorical([1.0, 3.0, 5.0, 7.0, 9.0]), # sigma for blue
    Categorical([1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0]), # contrast_factor for red
    Categorical([1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0]), # contrast_factor for green
    Categorical([1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0])  # contrast_factor for blue
]
```

Thank you