Machine Learning Methods for Ship Detection in Satellite Images

Group 2: Yifan Li, Huadong Zhang, Xiaoshi Li, Qianfeng Guo
{yil150, huz095, qig020, xil758}@ucsd.edu

Abstract:

In this project, we compared the performances of several machine learning methods on binary classification task, then the results were improved by HOG feature extraction in data preprocessing. Furthermore, we implemented convolutional neural network (CNN) and reached an accuracy of 99%. Based on the CNN model, we accurately detected all ships in satellite images of San Francisco Bay Area with bounding boxes using sliding windows detection algorithm.

Data:

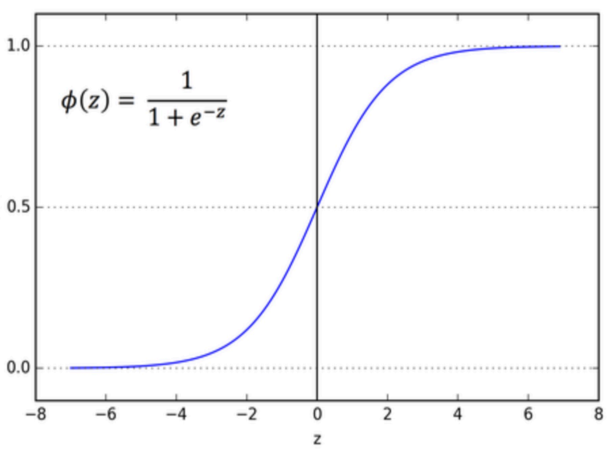
1. For classification:
   - 2800 RGB images (80*80)
   - Ship: 700 No ship: 2100
2. For detection: 1777*2825 RGB images

Methods:

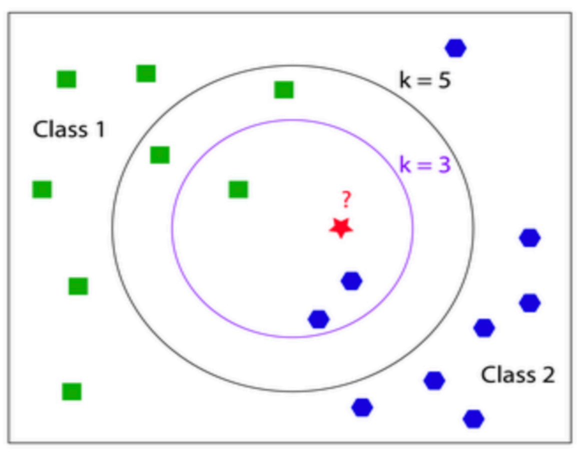
1. Image Classification

1.1 Machine Learning Methods

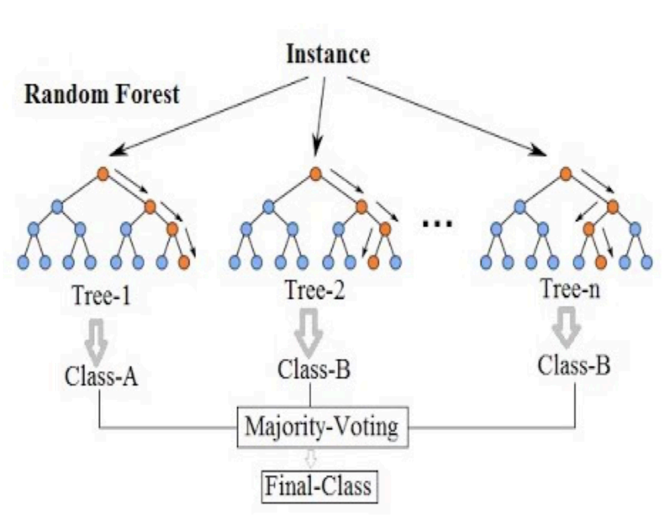
- LR



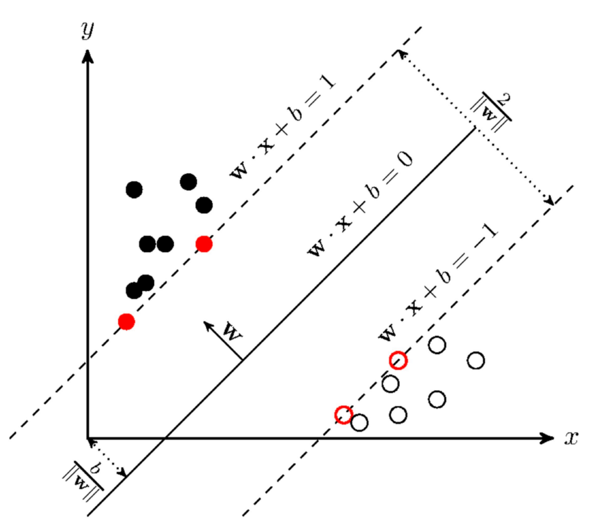
- k-NN



- RF



- SVM



- LR: Logistic Regression
- K-NN: k-Nearest Neighbors
- RF: Random Forest
- SVM: Support Vector Machine

1.2 Convolutional Neural Network

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 80, 80, 32)	896
max_pooling2d_1 (MaxPooling2	(None, 40, 40, 32)	0
dropout_1 (Dropout)	(None, 40, 40, 32)	0
conv2d_2 (Conv2D)	(None, 40, 40, 32)	9248
max_pooling2d_2 (MaxPooling2	(None, 20, 20, 32)	0
dropout_2 (Dropout)	(None, 20, 20, 32)	0
conv2d_3 (Conv2D)	(None, 20, 20, 32)	9248
max_pooling2d_3 (MaxPooling2	(None, 10, 10, 32)	0
dropout_3 (Dropout)	(None, 10, 10, 32)	0
conv2d_4 (Conv2D)	(None, 10, 10, 32)	102432
max_pooling2d_4 (MaxPooling2	(None, 5, 5, 32)	0
dropout_4 (Dropout)	(None, 5, 5, 32)	0
flatten_1 (Flatten)	(None, 800)	0
dense_1 (Dense)	(None, 512)	410112
dropout_5 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 2)	1026

Total params: 532,962
Trainable params: 532,962
Non-trainable params: 0

2. Sliding Windows Detection

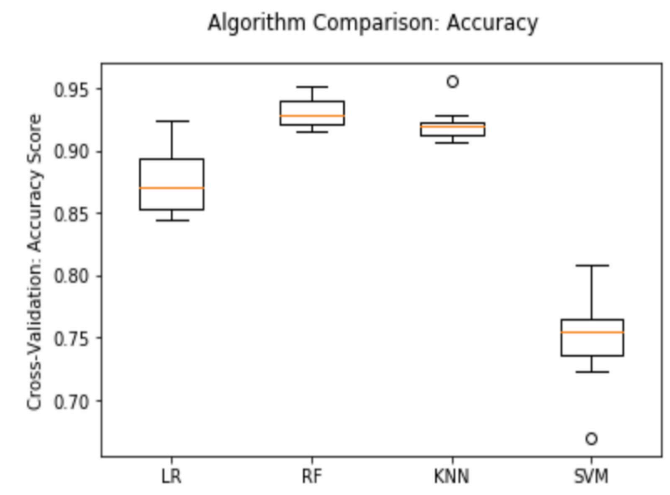
- Non-Maxima suppression

Results:

1. Machine Learning Methods Comparison

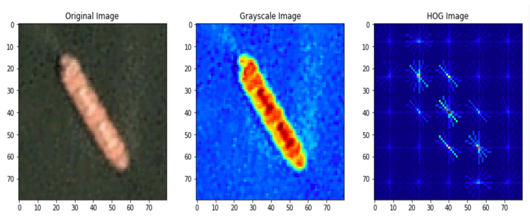
Preliminary Result

```
Compare Multiple Classifiers:

K-Fold Cross-Validation Accuracy:

LR: 0.875893 (0.025707)
RF: 0.930804 (0.011512)
KNN: 0.921429 (0.012969)
SVM: 0.750893 (0.035926)
```



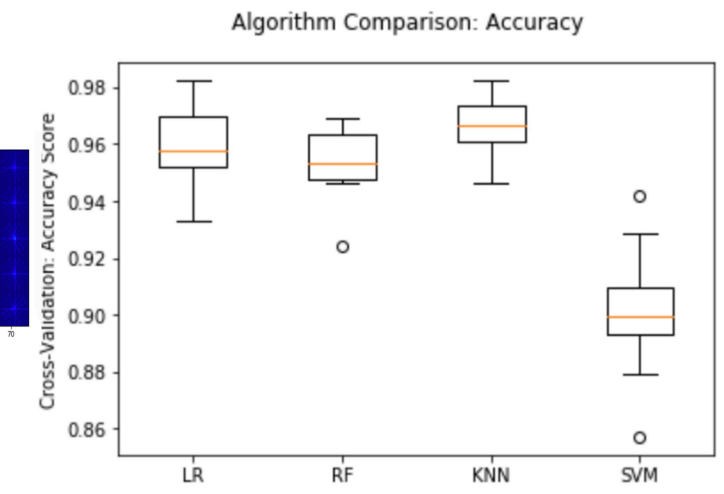

```
LR = Logistic Regression
RF = Random Forest
KNN = K-Nearest Neighbors
SVM = Support Vector Machine
Runtime: 1865.1310410499573s
```

HOG extraction



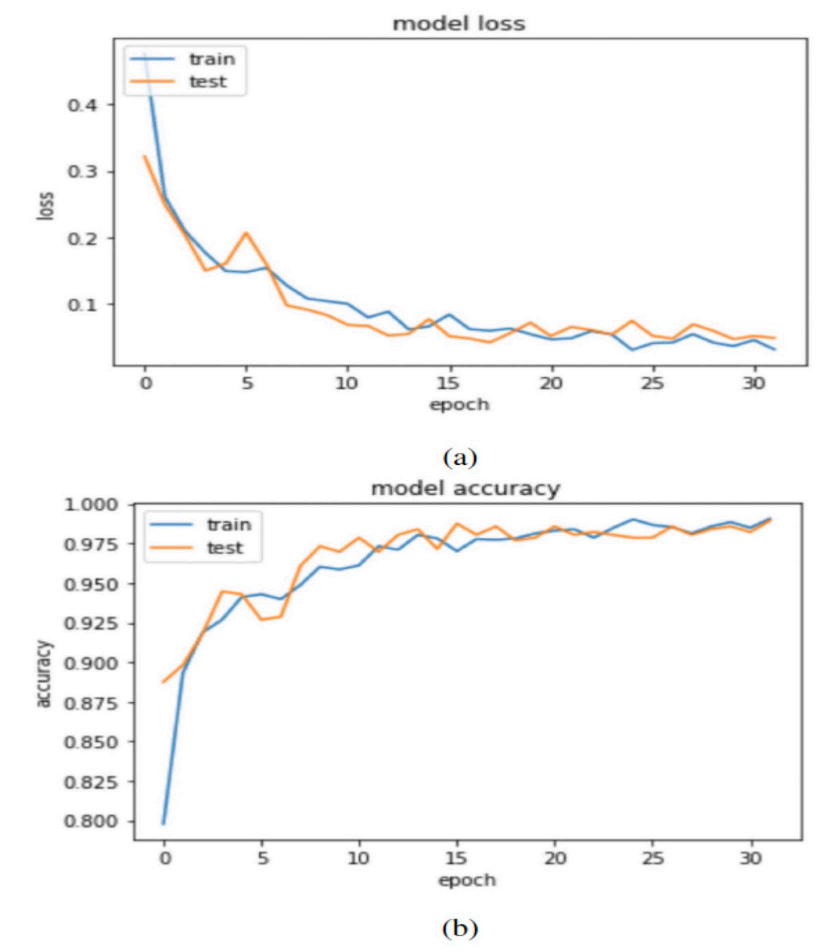
Improvement by HOG

```
Compare Multiple Classifiers:

K-Fold Cross-Validation Accuracy:

LR: 0.958929 (0.014644)
RF: 0.953571 (0.012658)
KNN: 0.966518 (0.010229)
SVM: 0.900893 (0.022746)
```




```
LR = Logistic Regression
RF = Random Forest
KNN = K-Nearest Neighbors
SVM = Support Vector Machine
Runtime: 27.46461319923401s
```

2. Convolutional Neural Network



- Optimizer: Adam
- Mini-batch size: 32
- Epochs: 32
- Accuracy: 99%

3. Sliding Windows Detection

- Window size: 80*80
- Stride: 10
- NMS threshold: 0.2

Conclusions:

- By HOG feature extraction in data preprocessing, traditional machine learning methods could reach over 90% accuracy.
- The superiority of convolutional neural network on image classification problem was proved again with 99% accuracy.
- Given a relatively small image to search, sliding windows detection is still a reliable algorithm for object detection task.

Discussion:

If given large images to search, sliding windows detection will be inefficient. In situations like this, we have to find a well-labeled dataset that includes the locations of objects to train CNN model with more advanced detection algorithms like YOLO.