Boat Recognition: A Comparative Study over the Image Classification Methods

Tianming Yang, Mengting Zhang, Minjiao Gu

***Abstract*—In the past decade, boat recognition technique has widely developed and been applied in many scenarios, especially military defense and maritime safety. However, most of current boat recognition researches focus on applying satellite vessel images. To fill in this gap, this paper discusses on recognizing visible boat pictures captured by cameras because there is some requirements on it. In this paper, visible images of boats containing five boat types are obtained from Google. After applying histogram of oriented gradient (HOG) to extract features and feature selection, some machine learning algorithms, including k-nearest neighbour (kNN), support vector machine (SVM), decision tree, AdaBoost and random forest (RF), are applied to do classification. The results demonstrates that Random Forest shows the best classifying performance with 75% accuracy on the given data-set compared with other traditional machine learning methods. To do the comparison, convolutional neaural networks (CNN) is applied and achieve approximately 96% accuracy.**

***Index Terms*—Boat Classification, Visible Boat Images, Image Recognition, Histogram of Oriented Gradient (HOG), Random Forests, Convolutional Neaural Networks (CNN)**

I. Introduction

Boat recognition is a technique to recognize boat types among different vessel types images,which is an application of image recognition. In the past decades, significant advancements have been made in image recognition study thanks to the development of machine learning as well as deep learning methods.

Based on different images collection approaches, current boat recognition study can be divided into two common types, study on satellite images captured from synthetic aperture radar (SAR) and that on visible images obtained from surveillance. Nowadays, boat recognition using satellite images is really mature since it has gone through wide-range development over ten years. However, there are rare researches on such technique based on visible images captured from cameras. But since there are more and more application scenarios in such technique, some researchers have developed research interest in such topic during the past two years.

There are two notable requirements on boat recognition using visible images. Firstly, it is meaningful for military defense. For coastal countries, maintaining the safety of coastline is an important mission for mainland defense because it can help resisting maritime illegal activities such as maritime terrorism, piracy, illegal immigration and trafficking of drugs and weapon [1]. However, it is not reasonable to monitor suspicious boats the whole day manually. Therefore, to address this problem, the European research project AMASS develops a unmanned surveillance platform that is equipped with high definition cameras to recognize suspicious vessels along the bay and on the ocean [1]. Moreover, it also has requirement on automatic seafaring vessels. Self-driving is now gradually applied on maritime transportation. In May 2018, researchers at MIT has developed a 3D printed self-driving boats to help reduce the road traffic in city [2]. In June 2018, a self-driving sailboat 'SB Met' developed by a Norway company called 'Offshore Sensing AS' has successfully crossed the Atlantic, which became the first robot boat to go across Atlantic in the world [3]. Such development proposes the researches on boat recognition as well. According to International marine traffic regulations, to make sure the safety of boat transportation and get rid of the risk of boat crashes, such self-driving boat is required to be equipped with machine vision systems to automatically detect and recognize the other boats nearby. Because of the need of military defense and maritime transportation, it is necessary to do some research on boat recognition.

This paper is structured as follows. Section II is literature review and some related work is introduced in this section. In Section III, collection of data-set and pre-processing method is given. Section IV first gives a detailed description on the classification methodology applied in this project and then shows the results. Section V gives the analysis of the results. Section VI draws a conclusion of the whole project.

II. Literature Review

Image recognition provides a way to recognize different types of objects by processing and analyzing the images using computers, which is a promising field of artificial intelligence (AI). The development of image recognition has been through three stages: optical character recognition (OCR), digital image process and recognition and object recognition. It now has generous application in many fields.

Image Recognition technology includes the following process: information acquisition, pre-processing, feature extraction and selection and classifier design. For image recognition problem, the acquired data is a group of figures with or without labels. Pre-processing is image process, which includes binarization, imagery smoothing and enhancement. In pattern recognition, it is required to choose the the most important features in feature space by transform the original data. The most common feature extraction method is principal component analysis (PCA) and linear discriminant analysis (LDA). Finally, classification is the key part of image recognition. Thus, it is necessary to choose classifier and its parameter appropriately based on different data-set. There are several classifying approaches in machine learning, such as k-nearest neighbour (kNN) and support vector machine (SVM).

Recently, with the development of deep learning methods like convolutional neaural networks (CNN), it is extensively applied in image recognition.

As one of the most promising applications of image recognition, boat recognition has witnessed great development in the past decade. Several related work in this topic have appeared gradually.

In [1], machine learning method like SVM is applied to classify the data-set for three classes: clutter, irrelevant objects and suspicious boats. The author creates a two-stages-classification, in which the first stage is used to separate the clutter from the original data-set while the second stage is a binary SVM classifier to distinguish whether the object is suspicious boats or not. The total classifying accuracy can reach around 97%. However, some disadvantages existed in this paper: such two-stages-classification is not applicable to multi-type classification problems as it would increase computation complexity. And increasing the class types may result in low precision as well. To solve such problem, paper [4] introduces a boat detection model based on cNN to classify five types of boats, including pleasure boats, tugboats, houseboats, jet skis, and others (e.g., workboats and rowboats). The accuracy of such approach is only 70%.

In order to improve the accuracy, paper [5] applied CNN to classify the boats and used transfer learning method (ImageNet) to fine-tune the model. Its accuracy can reach 78% among 26 types of boats.

Based on the above researches, this paper studies image recognition process and apply several traditional machine learning methods to classify five-type boat image data-set.

III. Description of Methods Selected

Before classification, some process including data collection, pre-processing, feature extraction and selection, are needed. This section shows the approaches selected during these process in our project.

A. Data Collection

In this experiment, the dataset is from Kaggle. Several types of boats are provided in this dataset, such as cruise ship, kayak, ferry boat, sail boat and gondola. However, some type of boat in the dataset is not enough to support the experiment. For example, sailboat only has 30 images. As a result, we scraping more image from the internet. We also cleaned up some low quality images.

B. Pre-processing

1) Image Normalization: We start with image normalization for the image in order that all image in same standard. All images are rescaled to 512*512 in order to make all image in a same size. After that, some useless information should be removed. For boat type recognition, it will more focus on the physical shape feature, so we can considered that color can be remove. Binary threshold is used to transform image from color to binary color image.

Normalization can reduce the effects when the shape of boat changes. It makes a linear transform for the original

Fig. 1. Boat Types

image, and helping the computer to generate feature. It as easy for human eyes to detect object even though their color is different. However, for computer, image may be considered as a total different image if the color is a little bit different. So, it is important to make normalization in order to get a accurate result.

Fig. 2. Binary Image of Boat

As we can see from figure 1, it keep most of the information from original image, but remove the color. The image shows not only the shape of boat, but also shows the intensity of image. This process could help generating features from image using method such as hog, LPB, SIFT.

Whats more, the range of pixel intensity was change during the pre-processing in order to avoid the influence of high frequency noise and low frequency noise.

C. Image Enhancement

The quality of images is not same in our dataset. Some bad quality images exist in the dataset. For example, some photo is taken during night so it is hard to detect object because it is hard to distinguish between boat and background. Laplace filter is used here to enhance the image. It can sharpen image and make the boat easy to detect. In addition, it can also adjust the contrast ratio, darkness and color. Laplace transform equation:

$$\begin{aligned}\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + ... \\ ... + f(x,y+1) + f(x,y-1) - 4f(x,y),\end{aligned} \tag{1}$$

sharpen equation:

$$g(x,y) = f(x,y) + c[\nabla^2 f(x,y)], \tag{2}$$

where g is output, f is original image, c is coefficient and triangle is weight.

By applying this equation, we can get use Laplace transform to enhance the image.

Fig. 3. Sharpen Image

As the Fig. 3 shows, the image is sharpened by Laplace filter. The edge is more clear after being sharpened. It is easy to see the difference between background and boat.

In addition, we also test image enhancement methods such Sobel and Canny. Here is the image after Sobel and Canny enhancement. We can see from these images, they keep most of the information and detect the edge very well. However, comparing with Laplacian filter, it has more useless information such as water and building. As a result, Laplacian transform is selected as our enhancement method.

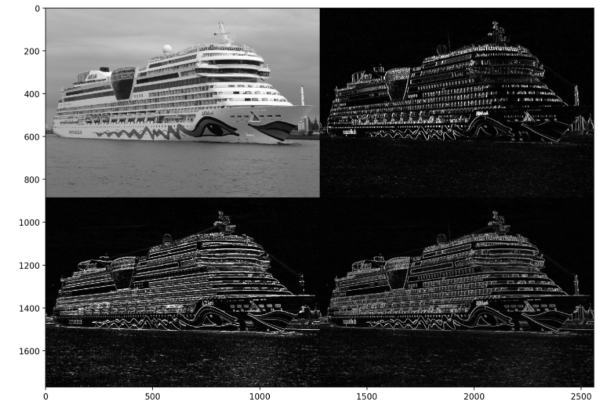

Fig. 4. Edge Detection

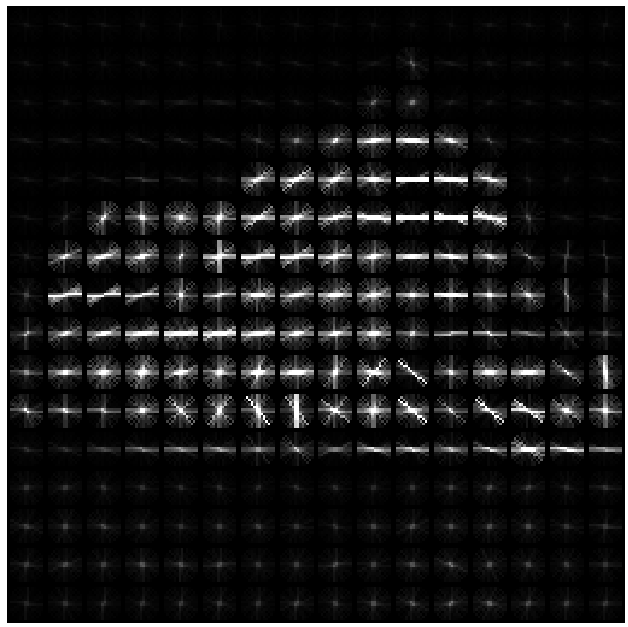

Fig. 5. HOG Feature Image

D. Feature Generation

1) Feature Extraction: For feature generation, appearance and shape is the most important thing that we considered in this project. By researching, we know that histogram of oriented gradient (HOG) performs well in this part. It is called histogram of oriented gradient. HOG can show the distribution of intensity gradients or edge directions very well. Images are dividing in to small cells and the histogram of gradient directions is calculated according to the pixels in cells. It counts the occurrences of gradient orientation in localized portions. Comparing with other methods, it has better performance when the geometric and photometric change happened. Considering our project, boat image can be taken in any photometric environment and make change the geometric shape while it is in the ocean. As a result, HOG is really suitable for the boat reorganization.

By applying HOG method to compute feature, 9604 features were generated.

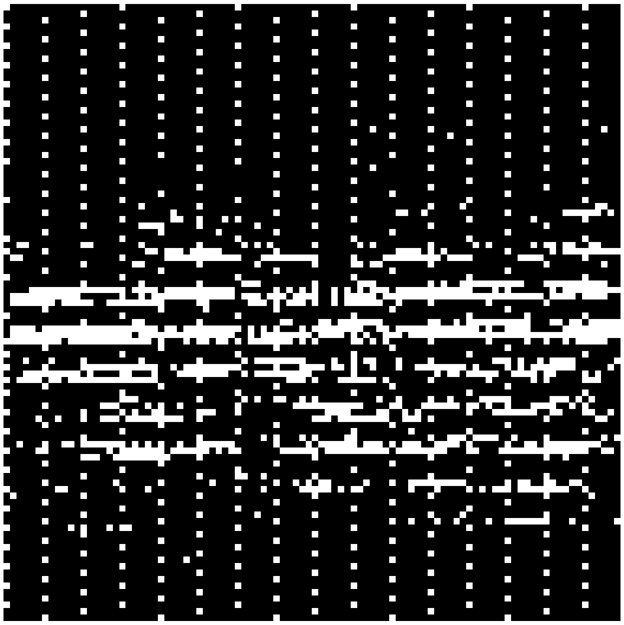

Fig. 6. 1500 Selected Features

2) *Feature Selection:* 9604 features are too much for the experiment. It contains lots of useless and noise information. For example, water and sky may consider as part of feature in those data. As a result, we decided to add feature selection in order to reduce some useless features. It can also help us to reduce dimension and increase the computing speed.

We use select K best features for the feature selection where K is the amount of feature that we wan to keep. Chi-squared Test was used to be our standard to evaluate the feature. Chi-squared Test can determine if there is a huge difference between observed frequencies and expected frequencies in categories. If the difference is huge, we can can consider that the category in dependent with another one. If the difference is small, we can consider that the category in independent with others. By applying Chi-squared Test, we can determine the dependency between each feature and select K best features based on it. Here is the equation to calculate chi-square Test value.

$$X^2(t,c) = \sum_{et\in(0,1)} \sum_{ec\in(0,1)} \frac{(N_{etec} - E_{etec})^2}{E_{etec}} \quad (3)$$

By testing different K value with different model, we select 1500 as our best feature in order to keep a better performance. Figure 6 is the selected feature on a 512*512 images. We can see most of features are located in the middle of image. That is because for most of images, boats are located in the middle of image. For those images, this feature selection can easily get the feature it wants. However, it caused some problem because not all boat in image are located in the middle part and some of them are portion of the boat. As figure 7 shows, it only contains part of cruise boat. If we applying feature selection on this image, not enough information can be provided for the classifier.

Fig. 7. Portion of Cruise Boat

TABLE I
PARAMETER SELECTION OF DIFFERENT MODELS

Model	Parameter
SVM	C=1
kNN	n_neighbors=11
DT	min_samples_split=20, min_samples_leaf=5
RF	n_estimators=100, max_depth=20
Adaboost	n_estimators=200, learning_rate=0.8

IV. CLASSIFICATION ON DIFFERENT MODELS

This section studies on several traditional classifiers, including SVM, Decision Tree, kNN, AdaBoost and Random Forests.

A. Parameter Selection

TableI shows the parameter selection of these models. We try different parameters, and choose the best ones after evaluation. In the statistical learning framework, Error = Bias + Variance. The prediction error consists of two parts, that Bias which is caused by simplicity of the model, and Variance is caused by the larger space for change due to the complexity of the model.

If we want to reduce the Bias of the model, it would increase the Variance of the model, and vice versa. If we believe in the authenticity of the training data and ignore the prior knowledge of the model, we will ensure the accuracy of the model on the training samples, which can reduce the Bias of the model, but this will make the model's generalization ability insufficient, resulting in overfitting, reducing the performance of the model on real data, increasing the uncertainty of the model. On the other hand, if we pay more attention to the prior knowledge of the model, adding more restrictions to the model in the process of learning the model can reduce the variance of the model and improve the stability of the model, but it will also increase the Bias of the model.

B. Model Evaluation

1) *Classification Report:* True Positive (TP): The real category is a positive example and the forecast category is a

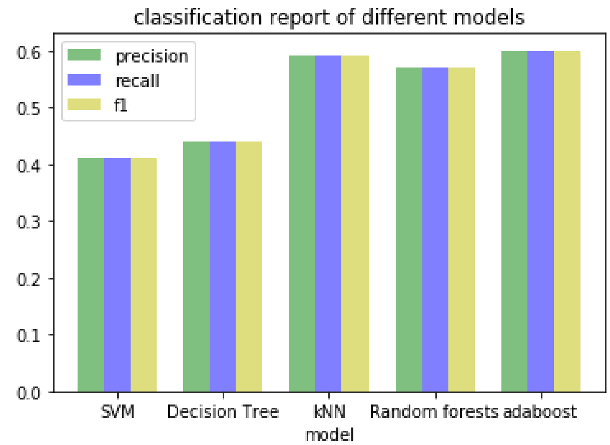

Fig. 8. classification report of different models

TABLE II
ACCURACY OF MODELS BASED ON 10-TIMES 10-FOLD CROSS VALIDATION

Model	SVM	kNN	DT	RF	adaboost
Accuracy	0.4251	0.5373	0.4111	0.5752	0.5875

positive example.
False Positive (FP): The real category is a negative example and the forecast category is a positive example.
False Negative (FN): The real category is a positive example and the forecast category is a negative example.
True Negative (TN): The real category is negative and the forecast category is negative.
Precision(P): $p = \frac{TP}{TP+FP}$
Recall(R): $R = \frac{TP}{TP+FN}$
F1: $F1 = \frac{2 \times P \times R}{P+R}$

Figure8 shows the precision, recall and f1-score of different classification.

2) Confusion Matrix: Confusion matrix can evaluate the accuracy of a classification. By definition a confusion matrix C is such that $C_{i,j}$ is equal to the number of observations known to be in group i but predicted to be j in group. Figure 9-13 show the confusion matrix of different models. SVM is the worst performer in all classification methods, where all test images are divided into sailboat. The sailboat is classified more correctly than others, but there are many other boats that are misclassified into sailboat. So, its precision is lower than its recall. This may be because sailboat has a larger proportion in the data-set, so it can be easier to classify an image into sailboat. Cruise ship and kayak are the second best, and their precision is higher than recall, which is the opposite of sailboat. ferry boat shows worst performance in knn,random forests and adaboost and perform a little better in decision tree, however gondola shows worst performance in decision tree which is the opposite of ferry boat.

C. Results and Analysis

To evaluate the performance of different models that we choose, we execute training based on the 10-times-10-fold

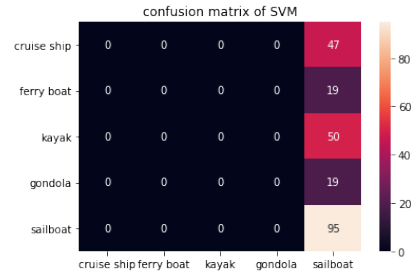

Fig. 9. Confusion Matrix of SVM

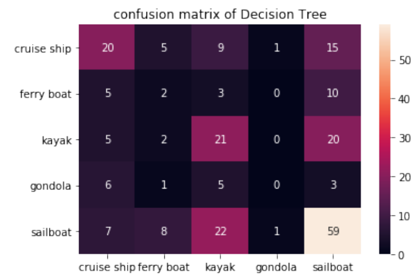

Fig. 10. Confusion Matrix of Decision Tree

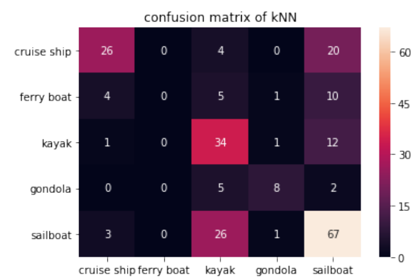

Fig. 11. Confusion Matrix of kNN

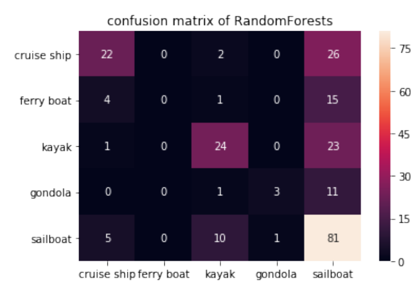

Fig. 12. Confusion Matrix of Random Forests

cross validation approach. The result is shown in tableII. SVM and decision tree have the lowest accuracy, and kNN performs much better than them. Random Forests and adaboost(DT) which are both based on decision tree have the best performance.
A random forest consists of multiple single trees, each based on a random sample of training data. They are usually more accurate than a single decision tree. Decision tree needs pruning to avoid overfitting. However, The feature set selected by each tree in the random forest is different, so the overfitting problem can be avoided. The tree of the random forest grows

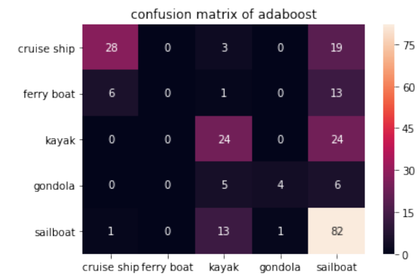

Fig. 13. Confusion Matrix of Adaboost

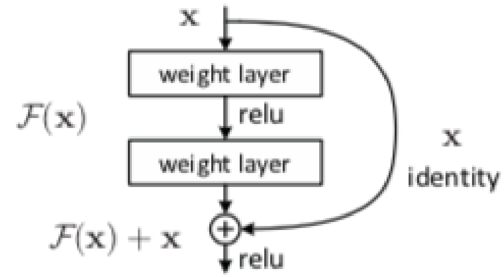

Fig. 14. Residual learning: a building block

completely and is not pruned, so naturally, the feature space is divided into more and smaller areas. Random forest learns every tree in the forests, and at each node, a random feature set for splitting is considered. Both mechanisms create diversity between trees in Random Forests, so that Random Forests show much better than decision tree.

Adaboost learns multiple classifiers by changing the weight of the training samples, and linearly combines these classifiers to improve the accuracy of the classification. In each iteration, the algorithm increases the weights of the samples that were misclassified before. And through weighted majority vote, the algorithm increases the weight of weak classifier with small classification error.

We choose decision tree as the weak classifier of adaboost. We can compare between Random Forests and adaboost which are both based on trees. Although adaboost shows a litter higher accuracy than Random Forests, the cost of adaboost is much more than Random Forests. So, Random Forests has better overall performance.

V. Classification based on CNN

A. *CNNs in image classification*

The Convolutional Neural Networks(CNN) has a good performance in image classification since AlexNet [6] used CNN in large-scale image classification tasks and the error rate was successfully reduced to 16.4%, which was about 10% lower than the second one in the ImageNet Large Scale Visual Recognition Challenge(ILSVRC) 2012. In ILSVRC 2014, Google Net [7] had made a significant breakthrough and won the top prize with a good error rate of 6.7%, which is half the price of the previous best result.

In ILSVRC 2015, Residual Networks(ResNet) [8]proposed by Microsoft Research Asia team won the image classification championship with an absolute advantage of 3.57% error rate. With the increase of network depth, the difficulty of training is increasing, and the accuracy of image recognition and classification is saturated or even begins to decline. The team put forward the idea of residual learning, that is, when the effect of network training can not be further, the residual function with learning value of 0 at the network level is easier to achieve better results than the identity function. Shortcut connection method is applied to link between some layers in the network to achieve residual learning, so that the appropriate accuracy will not decrease with the increase of the number of layers in the network.

In figure14, the formulation of F (x) + x can be realized by feedforward neural networks with shortcut connections. A shortcut connection is a connection that skips one or more layers. In ResNet, Shortcut connections simply perform identity mapping and their output is added to the output of the stacked layer [**?**].

So, we choose ResNet to do the boat type classification as comparison with traditional machine learning methods.

B. *Model design and implement*

The dataset are resized to 128x128 pixel size with 80% training set and 20% validation set. All images were normalized to ImageNet standards.

Fast.ai is a widely used deep learning framework built on PyTorch, which is simple and convenient for use. We use Fast.ai framework to train our model which is made use of CNN. We have adopted a strategy called transfer learning that uses existing networks trained for some tasks and reuses them for similar tasks. In our model, we use ResNet34 and weight that were previously trained on ImageNet.

C. *Results and Analysis*

We trained the model on the training dataset for 19 epochs (4 cycles). The results were as follows:
train_loss: 0.097
valid_loss:0.131
accuracy:0.964

During training, the train_loss in Figure16 and valid_loss are decline and accuracy increases15. Figure17 shows the confusion matrix of CNN. Almost all validation images are classified correctly, which is totally different from any other models we used above. Compared with traditional image classification methods, convolutional neural network has the ability of feature extraction, self-learning. By means of weight sharing, the number of neurons needed in the full connection layer is greatly reduced, the network structure is simplified, and the computational load required is significantly reduced. In addition, the convolutional neural network has the ability of learning and transferring. The trained network can apply the previously learned features to a new image classification task, thus effectively improving the poor generality of traditional image classification methods, and greatly improving the accuracy and efficiency of image classification.

However, there are still some problems that have not been solved well. The understanding of the mechanism of image

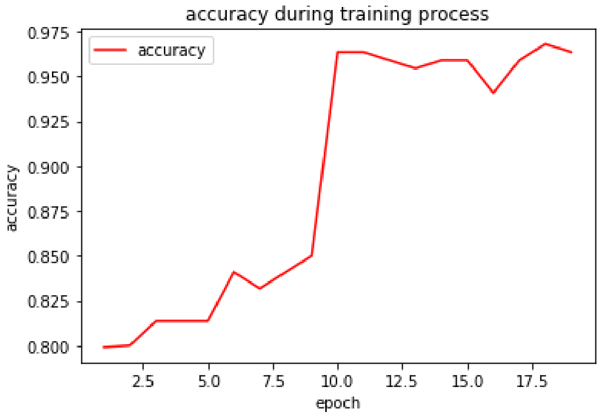

Fig. 15. accuracy during training process

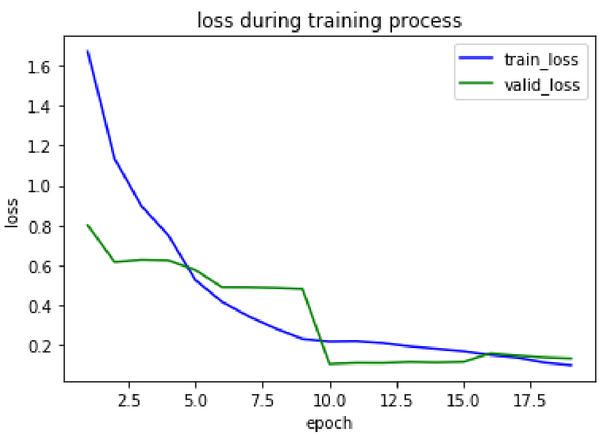

Fig. 16. loss during during training process

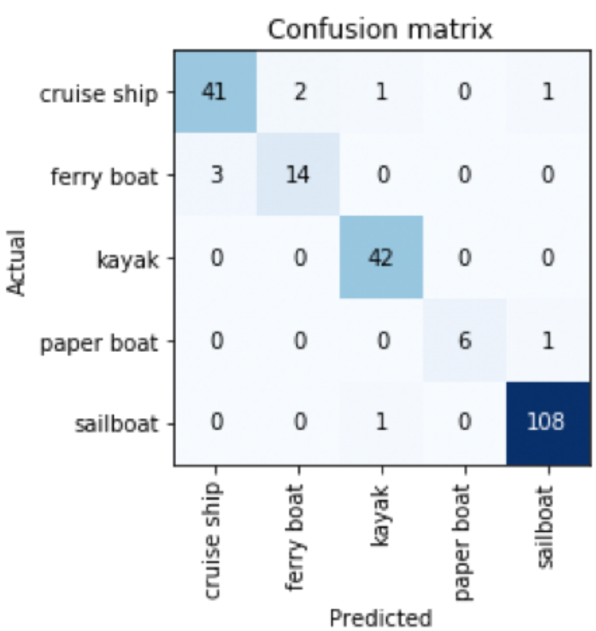

Fig. 17. Confusion matrix of CNN

feature extraction and classification is still not thorough, which leads to the need for some experience in setting network structure and network parameters. And with the deepening of network level, network degradation and over-fitting are prone to occur.

VI. Conclusion

This paper attempted different approaches of image classification methods to the problem of boat recognition in a data-set contains five different kinds of boat.

We did image pre-processing with image normalization and enhancement. And conduct feature extraction using HOG and feature Selection to select the best 1000 features from 9604 features.

We did experiments to evaluate the performance of image classification of some traditional machine learning methods, which are SVM, decision tree, kNN, adaboost and Random Forests. In this task, Random Forests showed the best performance with about 58% accuracy both in accuracy and training efficiency. We also find that combining weak classifiers to form a powerful classifier can significantly improve the performance of classification. Adaboost(DT) and Random Forests show higher accuracy than Decision Tree. And the data-set has huge impact on training, especially for these traditional machine learning algorithms.

For comparison, we trained a CNN model based on ResNet34. The performance of CNN is much better than other methods that we used with over 96% accuracy.

Acknowledgment

We would like to express the deepest appreciation Professor Mohammad Shafiee and University of Waterloo to make us have this opportunity practicing machine learning related skills during this project. We would also like to thank to TA Andrew Hryniowski and Tahereh Kamali for their contribution of ideas, guidance throughout our study.

References

[1] M. Teutsch and W. Kruger, "Classification of small boats in infrared images for maritime surveillance," in *2010 International WaterSide Security Conference*, pp. 1–7, IEEE, 2010.
[2] CNBC, "Mit designed a fleet of autonomous boats that could help reduce car traffic." https://www.cnbc.com/video/2018/05/31/mits-self-driving-3d-printed-boats-could-reduce-car-traffic.html, 2018.
[3] "Sb met, first robot vessel to cross the atlantic or is it?." http://www.oldsaltblog.com/2018/10/sb-met-first-robot-vessel-to-cross-the-atlantic-or-is-it/, 2018.
[4] J. K. Tatsuhiro Akiyama, Yosuke Kobayashi and K. Muto, "Cnn-based boat detection model for alert system using surveillance video camera," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pp. 669–670, IEEE, 2018.
[5] R. T. Maxime Leclerc, "Ship classification using deep learning techniques for maritime target tracking," in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 737–744, 2018.
[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012.
[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.
[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.