SHIP DETECTION AND CLASSIFICATION FOR KOMPSAT OPTICAL IMAGES USING DEEP LEARNING*

Jae Young Chang[1], Kwang Jae Lee, and Sun-Gu Lee
[1]Korea Aerospace Research Institute, 169-84, Gwahak-ro, Yuseong-Gu, Daejeon 34133, Korea,
Email: jychang@kari.re.kr

KEY WORDS: Machine learning, object detection, pose estimation, AIS, and high-resolution satellite image

ABSTRACT: Many studies have been conducted to apply deep learning-based technology to high-resolution satellite image analysis, and the key is how to efficiently apply models appropriate to our own problem. Our mission is to find and classify ships in KOMPSAT (Korea Multi-Purpose SATellite) optical images which include KOMPAT-2 (GSD 1 m), KOMPSAT-3 (GSD 70 cm), and KOMPSAT-3A (GSD 55 cm). The same model was used for the different satellite images with different GSDs by unifying the resolution in the preprocessing step. This also makes it possible to use public datasets from other optical satellites such as xView and DOTA for ship detection. Our deep-learning based analysis consists of three steps: detection, pose estimation, and type classification. Pose estimation means measuring the heading and the size of the ship. State-of-art object detection models usually don't produce oriented bounding boxes. This is why we separate pose estimation from the detection. We used AIS data manually registered to satellite images to create the ground truth of ship types. So the training data is prepared from different information sources by different methods for each model, and that's another reason to divide the analysis into three steps. In this study, we used the FasterRCNN + ResNeSt model for detection, and the ResNeSt model for both pose estimation and classification. It has performed well enough to achieve our initial goal, and the method seems to be efficient to adopt state-of-art models for further performance enhancement.

1. INTRODUCTION

Ship detection using satellite imagery is necessary in order to find a ship whose signal is not transmitted intentionally or inevitably. In this study, not only the detection of ships but also classification of the ship types is required using optical images taken from KOMPSAT-2 (K2), KOMPSAT-3 (K3), and KOMPSAT-3A (K3A). A number of studies have been done for ship detection in aerial or satellite images. Typically they used a single dataset with complete information about objects, and aim for a single model training. For example there are such datasets like xView, DOTA, Airbus Ship Detection, and HSRC2016. Although this end to end approach is preferred by the deep learning community, it does not apply to situations where a complete dataset cannot be constructed due to insufficient information. For example, we used AIS information to identify ship types, but the signals are frequently missed when acquired from satellites.

Our task is not to invent the new designs of deep neural networks (DNN), but find cutting-edge technologies that fit our problems. For this reason we rejected networks which outputs oriented bounding boxes (OBB) which is not handled by most recent detection networks. As a result, we separated the algorithm into three stages for the completeness of training data and efficient adoption of state-of-art networks. Whereas we unified the GSD for images from different satellites to secure enough amount of training data.

2. PROPOSED METHOD

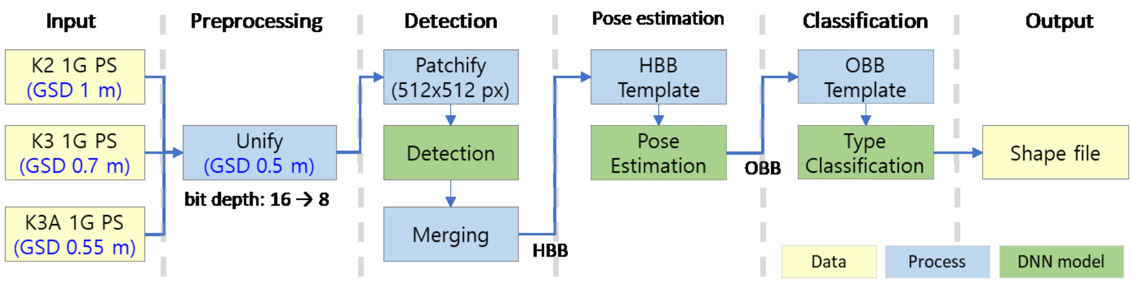


Fig. 1. Overall process of the proposed algorithm

* This study was supported by "Development of satellite based system on monitoring and predicting ship distribution in the contiguous zone" of the Korea Coast Guard.

The overall process of the proposed algorithm is shown in the Figure 1, there are three DNN models for detection, pose estimation, and classification.
The 1G pan sharpened standard product of KOMPSAT is used as the input. An input image is scaled to GSD 0.5 m and the bit depth is compressed from 16 to 8 bits in the 0.1% to 99.9% percentile range. And it is split into 512 × 512 pixel-sized patches by 256-pixel intervals. The contrast of the patch is normalized to fit N(128, 42) for an area of of 1 km^2 to minimize the environmental variation. They are fed into the object detection model to obtain the HBB of ships. Even after merging overlapped boxes via non max suppression (NMS), additional estimates of the pose and size are required. A patch is extracted with 25% margin from each HBB and scaled to 96 × 96 pixels. It is fed into the pose estimation model to obtain the ship's pose and size. In fact, the pose estimation model is a regression model that outputs four numbers indicating the end positions of the longitudinal and lateral axes of the ship. At this point we have the ship's oriented bounding box (OBB).
A patch is extracted from an OBB extended with a margin that is 40% of the breadth and resized to 96 × 96 pixels. The ship type is estimated by feeding the patch to the classification model. Finally, the position, pose, size, and type are summarized and exported as a shape file.

2.1 DNN models

FasterRCNN with FPN was chosen for the ship detection, and the base model is ResNeSt-101. The pose estimation and classification model is also the ResNeSt-101 but the loss is set differently. ResNeSt is an improved model of the ResNet by introducing the split and attention network to reflect cross channel information. Detailed parameters are shown in Table 1.

Table 1. Parameters of DNN models

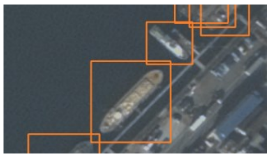
Objective	Model	Input	Output	Training loss	Validation Metric
Detection	FasterRCNN + FPN + ResNeSt-101	3 channels (RGB), 768 × 768 pixel	HBB	Huber loss (box), Softmax (class)	Mean average precision (mAP)
Pose estimation	ResNeSt-101	3 channels (RGB), 96 × 96 pixel	The end point coordinates of the longitudinal and lateral axes	Huber loss	Mean normalized correlation coefficient (mNCC)
Classification	ResNeSt-101	3 channels (RGB), 96 × 96 pixel	A class index (21 classes)	Softmax	Accuracy

2.2 Training and validation data

Detection

We prepared three types of datasets for detection. The first set is the xView whose images are taken from the WorldView-3. There are 10 classes in xView related to ships, but we've narrowed them down to a single class. The second set is the DOTA whose images are taken from various sources. The final set is newly created using KOMPSAT images. For all datasets, we correct the GSD to 50 cm and apply the contrast normalization as in the preprocessing step. Since our goal is to find and classify ships in KOMPSAT images, the validation set is selected only from the KOMPSAT dataset. Examples and label quantities are shown in Table 2.

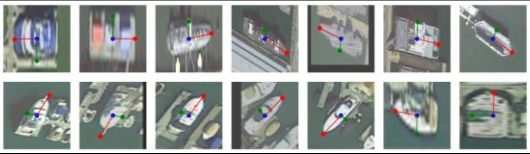
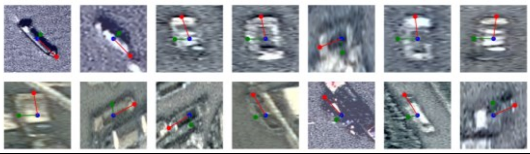
Table 2. Datasets for detection

Dataset	xView + DOTA	KOMPSAT
Example	 	
Quantity & Usage	31,243 labels Training 100%	8,762 labels Training 80%, Validation 20%

Pose estimation

We prepare two types of datasets for pose estimation. The first is from DOTA, and the second is from the KOMPSAT dataset. xView has been excluded because it has no pose information. A training patch is extracted from each OBB and resized to 96 × 96 pixel as in the preprocessing step. A label consists of two coordinates representing the endpoints of the longitudinal and lateral axes in normalized patch coordinate. As in the case of detection, validation samples are selected only from the KOMPSAT dataset. Some examples and quantities of each dataset are shown in Table 3. The red and green lines in the figures indicate the longitudinal and lateral axes.

Table 3. Datasets for pose estimation

Dataset	DOTA	KOMPSAT
Example		
Quantity & Usage	41,443 samples Training (100%)	12,123 samples Training (80%), Validation (20%)

Classification

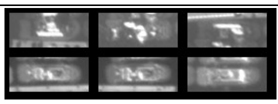
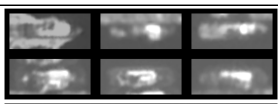
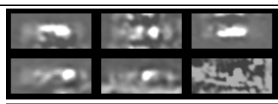
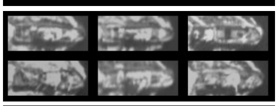
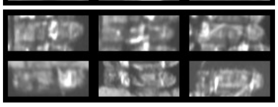
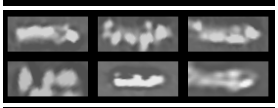
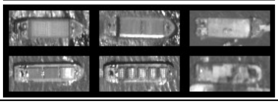
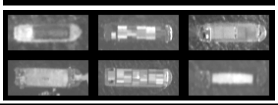
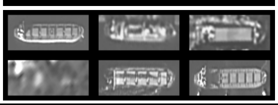
The exact ship type cannot be determined by visual inspection on satellite imagery. Even xView has labels of 10 types of ships, it cannot be reproduced in our KOMPSAT images. So we decided to obtain ship type information from AIS static data which is transmitted from ships. The number of ship types was reduced to 21 classes from AIS ship types, and the mapping table is shown in Table 4.

Table 4. AIS type to class ID mapping table

AIS type	Class (ID)	AIS type	Class (ID)	AIS type	Class (ID)
0 ~ 19	Not used	36	Sailing (6)	55	Law Enforcement (14)
38 ~ 39	Not used	37	Pleasure Craft (7)	56 ~ 57	Local Vessel (15)
20 ~ 29	WIG (0)	40 ~ 49	HSC (8)	58	Medical Transport (16)
30	Fishing (1)	50	Pilot Vessel (9)	59	Noncombatant ship (17)
31 ~ 32	Towing (2)	51	Rescue vessel (10)	60 ~ 69	Passenger (18)
33	Dredging (3)	52	Tug (11)	70 ~ 79	Cargo (19)
34	Diving ops (4)	53	Port Tender (12)	80 ~ 89	Tanker (20)
35	Military ops (5)	54	Anti-pollution (13)	90 ~ 99	Not used

The geometric errors between the satellite imagery and the AIS data were manually adjusted, but we admit some inconsistencies still exist. But at the same time, it'll be much more accurate information than a visual inspection. Assuming that the ship color and ship type are not related, the patch has only the luminance component. The dataset is created only from the KOMPSAT images, and some examples and quantities are shown in Table 5.

Table 5. Datasets for classification

Class	K3A	K3	K2
Tug			
Fishing			
Cargo			
Quantity & Usage	1,191 samples Training (80%), Validation (20%)		

3. RESULT AND DISCUSSION

Training results

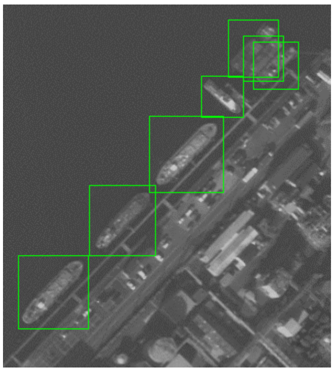
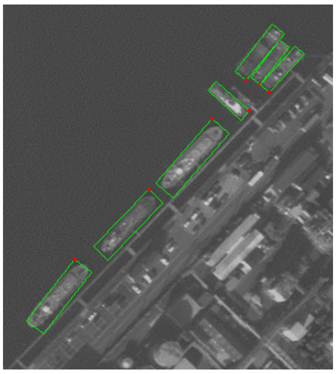
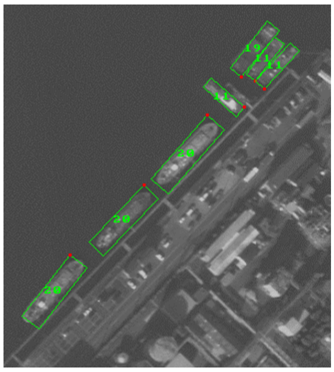
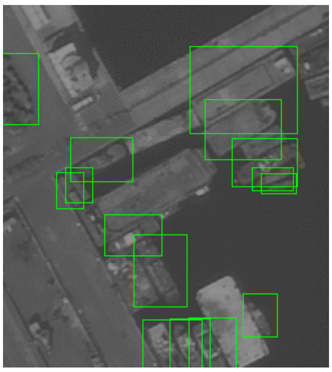
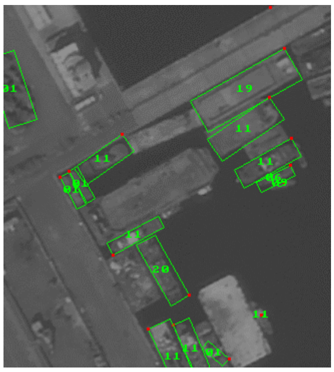
All results are summarized in Table 6.
For detection, there are three configurations of training. Training set A consists of KOMPSAT images, and training set B consists of xView and DOTA images. And we also have a mixture of A and B as a third configuration. The validation set consists of KOMPSAT images only and is same for all configurations. The best performance was mAP 77.1 achieved by training the set A.
For pose estimation, training set A consists of KOMPSAT images, and training set B consists of DOTA images. The validation set consists of KOMPSAT images only and is same for the three training configurations. The best performance was mNCC 0.678 achieved by training the set A.
For classification, there is only one training configuration with KOMPSAT images. The best performance was 75.7% accuracy.

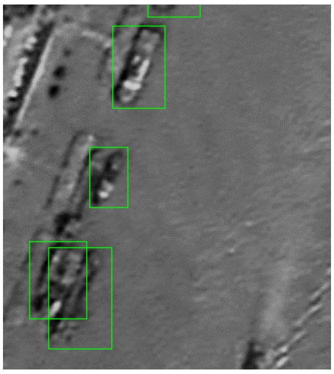
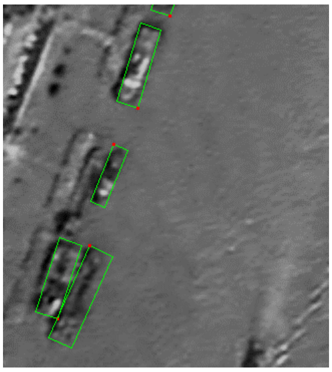
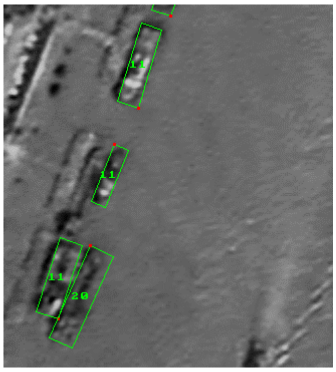
Table 6. Training results of models

Model (Training set)	Detection			Pose estimation			Classification
	A	B	A+B	A	B	A+B	
Epoch (best/total)	190/200	200/200	150/200	1450/1500	520/1500	830/1500	1500/3000
Validation metric (best)	**77.1**	76.8	76.5	**0.678**	0.643	0.673	**75.7**
Training loss (best)	3.15×10^{-2}	3.77×10^{-2}	3.68×10^{-2}	4.06×10^{-3}	7.60×10^{-3}	5.44×10^{-3}	2.21×10^{-3}

It can be observed that both detection and pose estimation achieve the best performance when the same image source (KOMPSAT) is used for training and validation. However, similar performance can also be achieved using training sets from other image sources. This means we can have a model before creating a training set for the target satellite and improve the productivity dramatically by reducing the amount of human visual inspection and manual annotation. Some examples of inference by the best parameters are depicted in Table 7.

Table 7. Step-by-step analysis result example

Image	Detection	Pose estimation	Classification
K3A			
K3			

K2			

4. CONCLUSION

This study proposes a three-step structure of ship detection and classification. It is designed to cope with training data incompleteness and efficiently adopt cutting-edge networks. Training data from other sources were also used and it shows that training samples from other sources have drawn comparable performance. So it can be used for effective production of the training data for the target.

To improve the performance of the proposed algorithm, we need to fine-tune the hyper parameters, the depth of the base network, and also try different models. And we should admit that this study did not take into account the smallest size of detectable ships, so even ships that are too small to be identified by the human eye are included in the training and validation set. Therefore, clarifying the target specification and refining the training data are our ongoing topic.

Also, since the amount of the classification training is relatively small, we continue to collect KOMPSAT images aligned with the AIS data.

5. REFERENCE

Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., and McCord, B., 2018. xView:Objects in context in overhead imagery. arXiv:1802.07856.

Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., and Zhang, L., 2018. DOTA: A largescale dataset for object detection in aerial images. CVPR, pp. 3974–3983.

Liu, Z., Yuan, L., Weng, L., and, Yang, Y., 2017. A High resolution optical satellite image dataset for ship recognition and some new baselines. ICPRAM, pp. 324-331.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K., 2017. Aggregated residual transformations for deep neural networks. arXiv:1611.05431.

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Muller, J., Manmatha, R., Li, M., and Smola, A., 2020, ResNeSt: Split-Attention Network. arXiv:2004.08955