


On Imbalance Problem in a Fully Automatic SAR Oil Spill Monitoring System

Kan ZENG, Ming-Xia HE

Ocean Remote Sensing Institute, Ocean University of China



表2 海面溢油监测结果细节

序号	海面溢油样本	中心经度	中心纬度	面积 (km²)
1		E 119°52'55"	N 38°45'25"	2.26
2		E 119°54'29"	N 38°43'26"	0.56
3		E 119°57'12"	N 38°41'42"	3.79
4		E 120°06'23"	N 38°38'42"	1.68
5		E 120°00'27"	N 38°39'16"	0.7
6		E 120°01'22"	N 38°39'57"	14.78
7		E 120°11'11"	N 38°38'45"	19.74
8		E 120°04'01"	N 38°35'01"	1.1
9		E 120°04'26"	N 38°36'39"	11.75
10		E 120°04'42"	N 38°32'08"	1.81
11		E 119°03'32"	N 38°26'38"	0.44

Fully automatic processing

Automatic generate and send report

compatible various SAR products

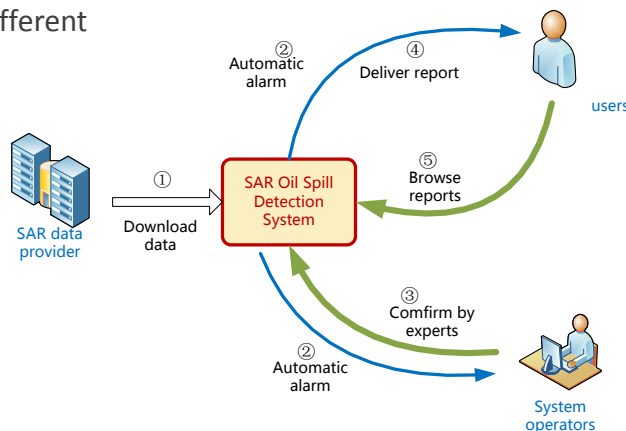
all algorithms are suitable to different SAR products.

Browser/Server Architecture

Easy to access system with web browser via internet

Fast delivery

less than 5 minutes.



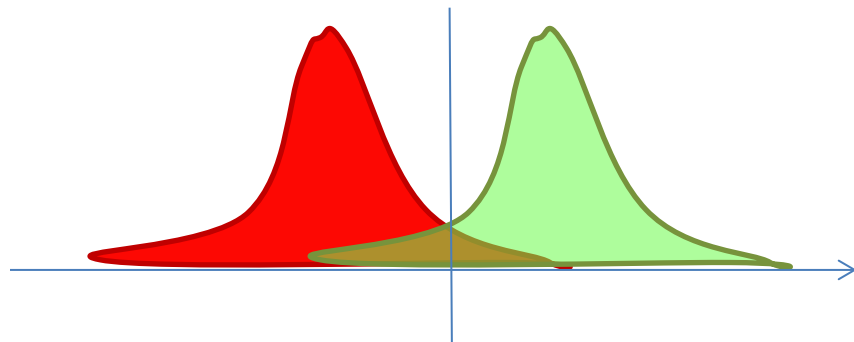
High classification performance

Detection rate $\geq 80\%$

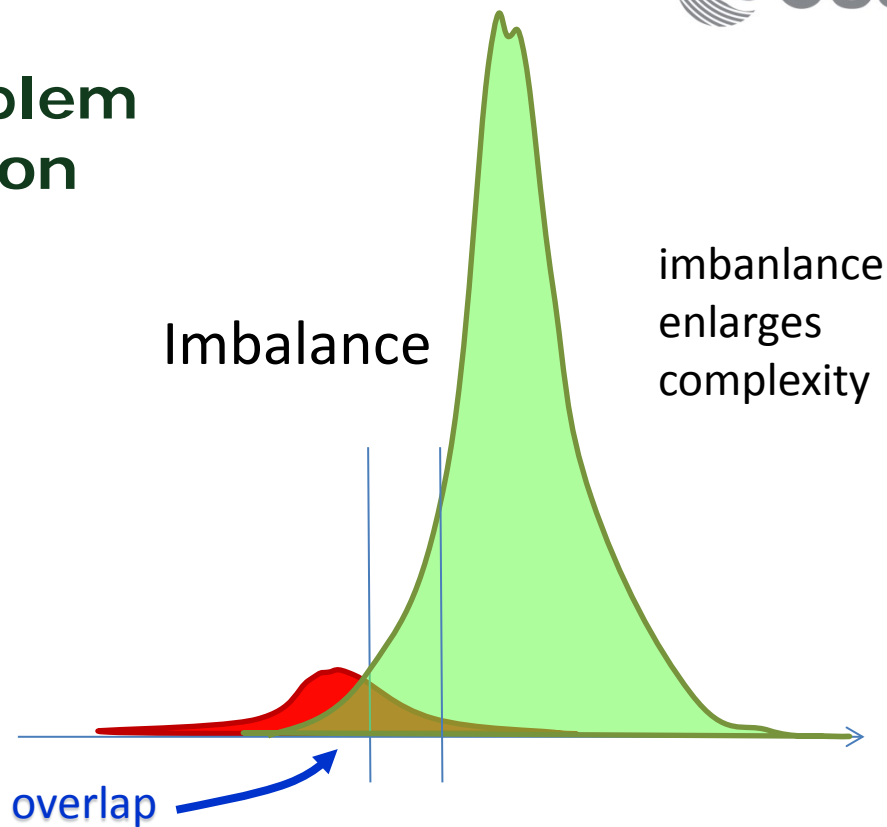
False Discovery rate $\leq 20\%$

Imbalance Problem of Classification

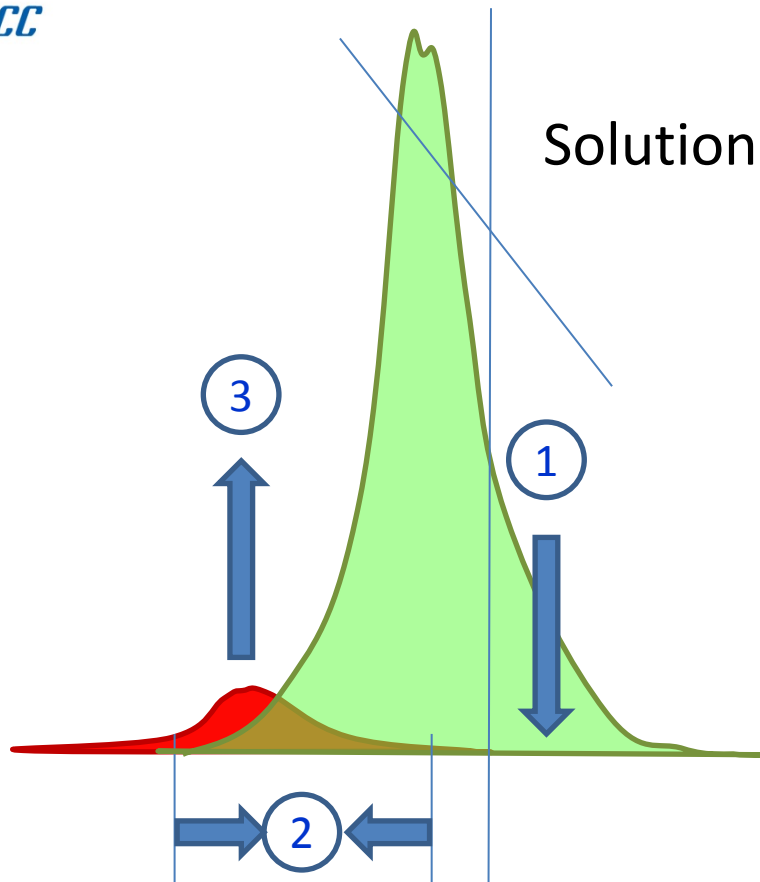
Balance



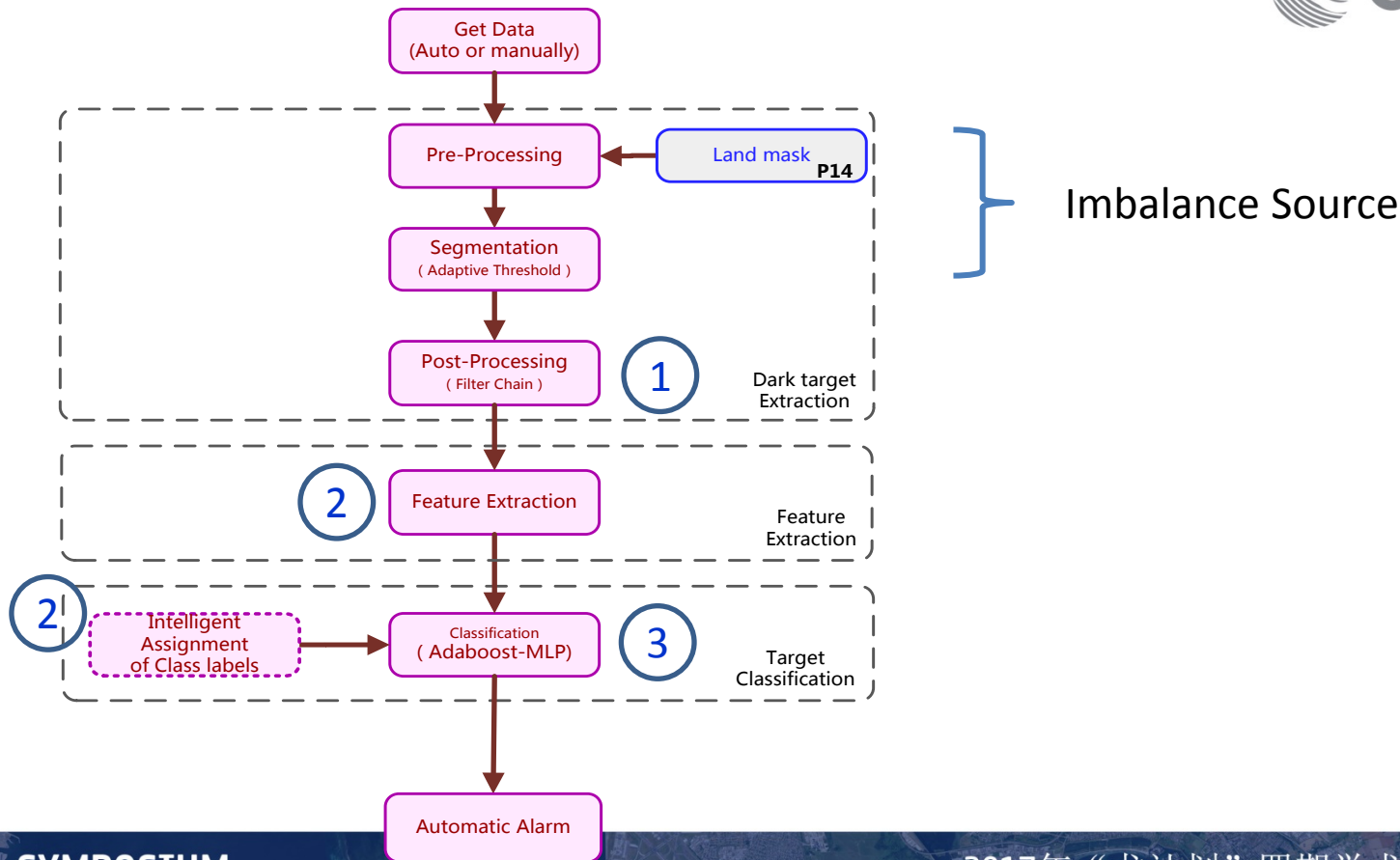
Imbalance



Solution for imbalance problem

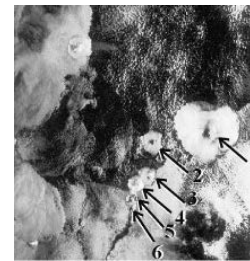
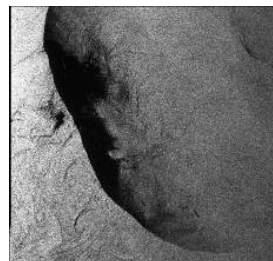
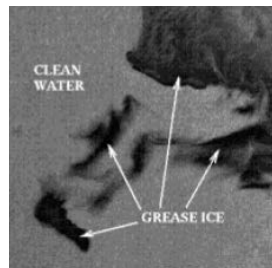
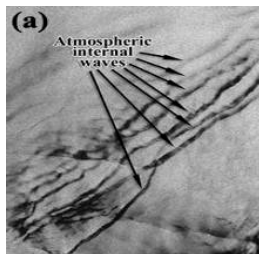
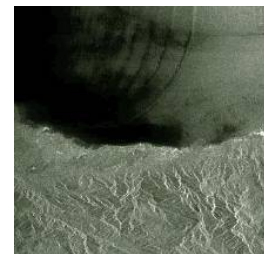
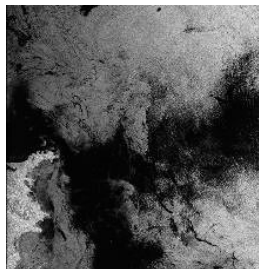


④ classifier insensitive to imbalance problem



Why imbalance ?

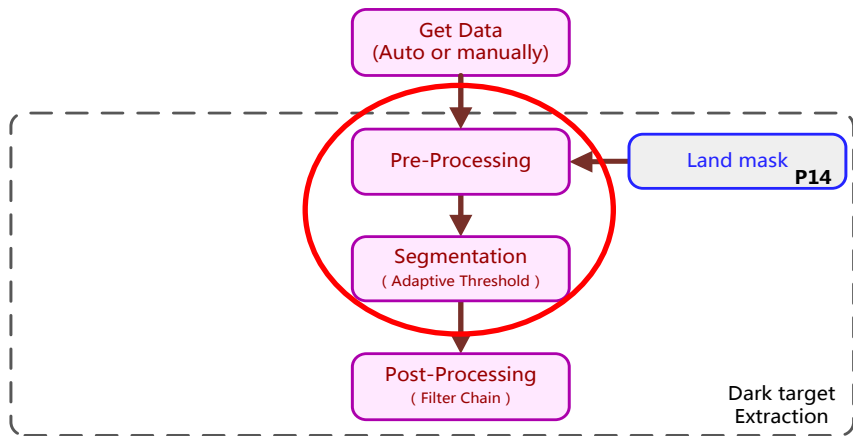
- The types of look alike are much more than that of oil spill.

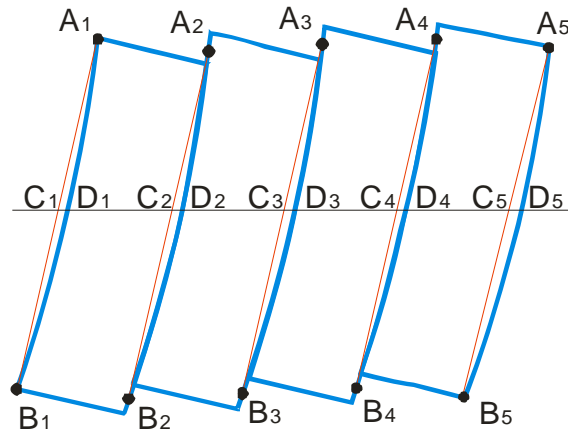
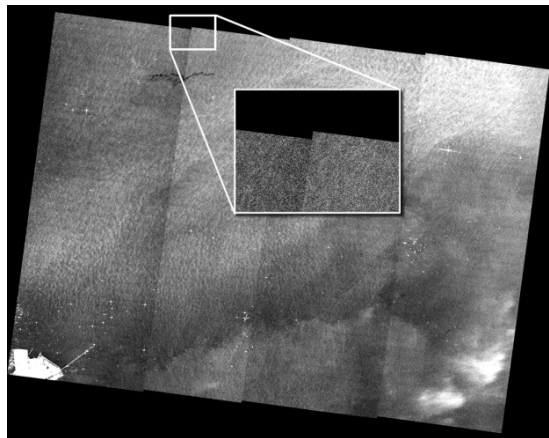


Why imbalance?

- Requirments of automation.

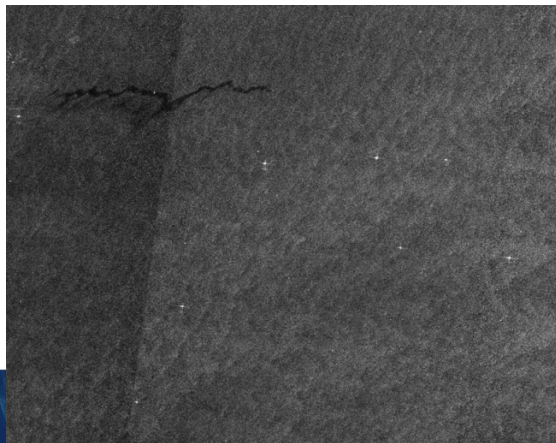
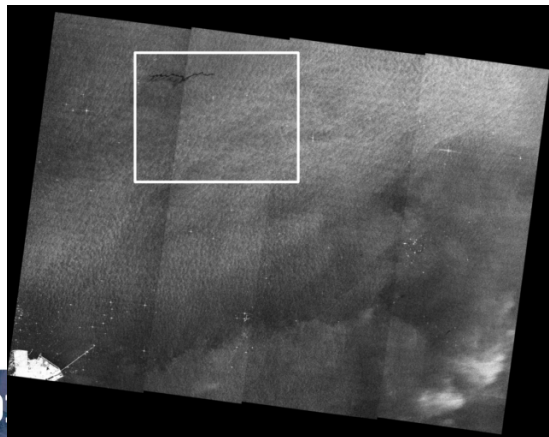
All targets extracted by segmentation algorithm must be processed by program.



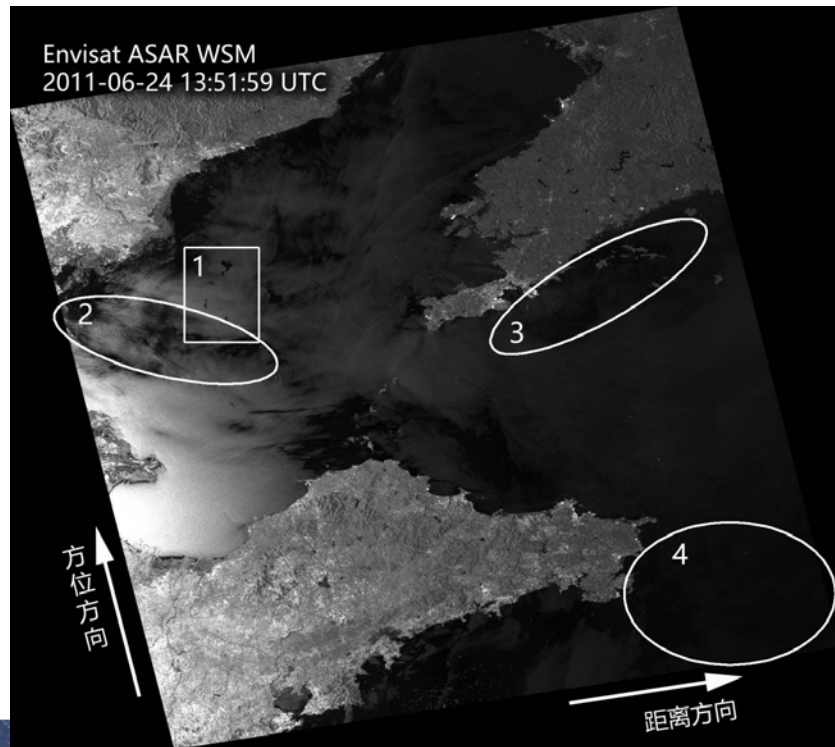
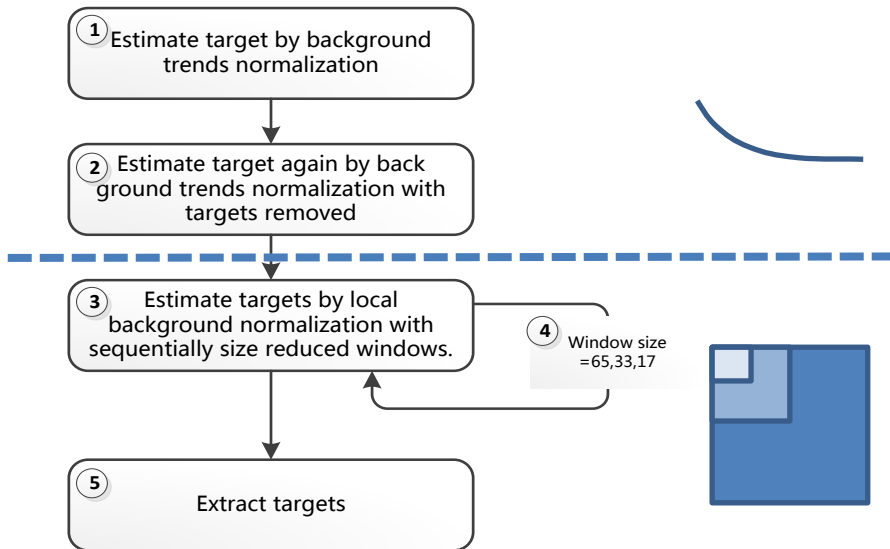


Remove image strip jump

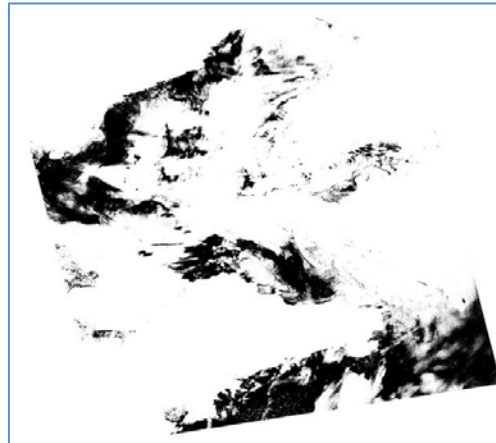
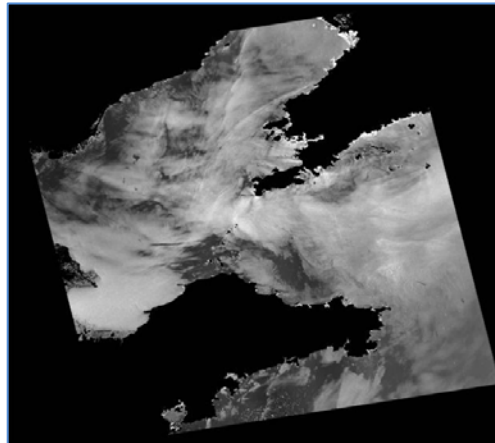
Comso Sky-Med Wide Region
Envisate WSM



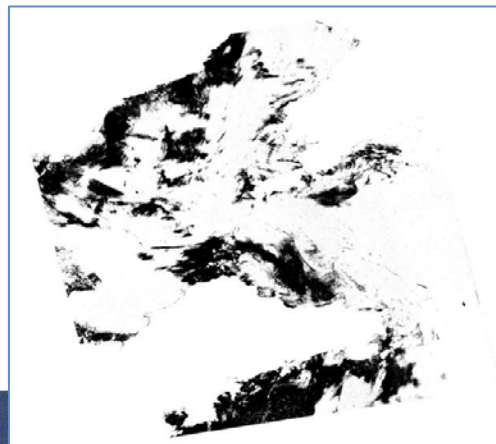
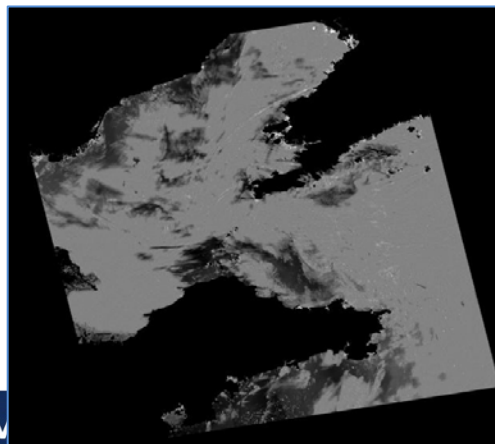
Adaptive threshold algorithm based on multi-scale background normalization



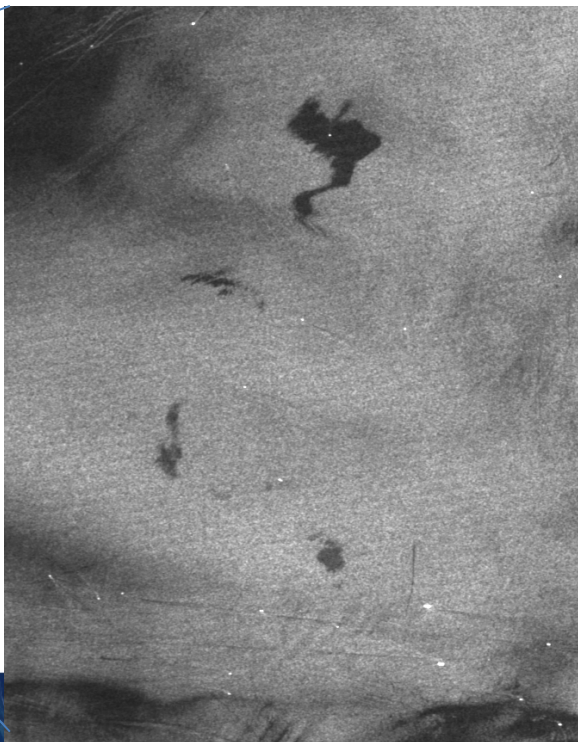
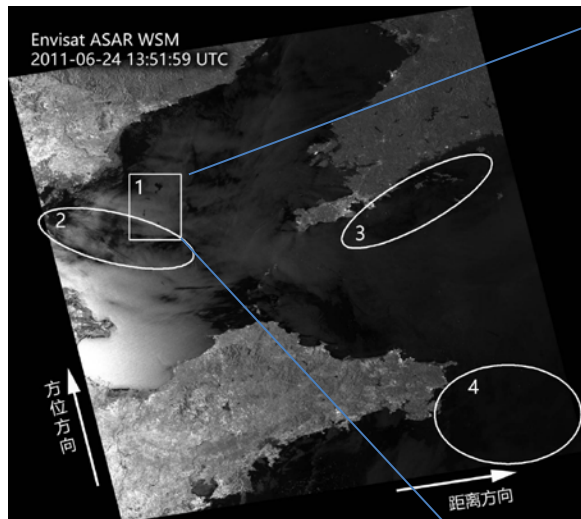
normalized
by background trend

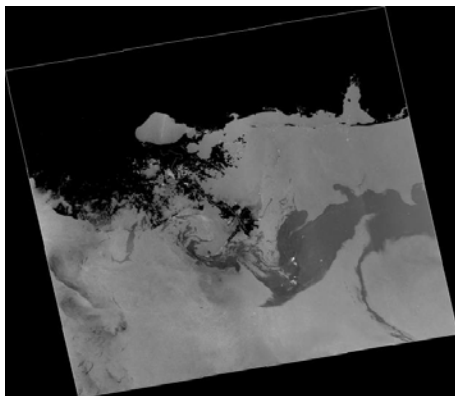


normalized
by local background

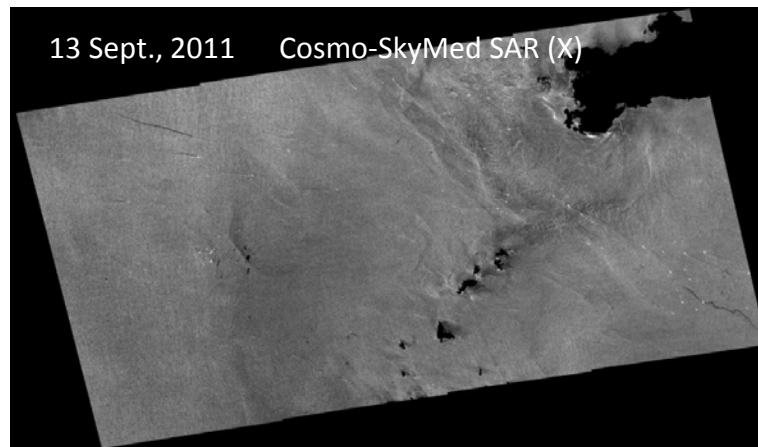


Adaptive threshold algorithm based on multi-scale background normalization

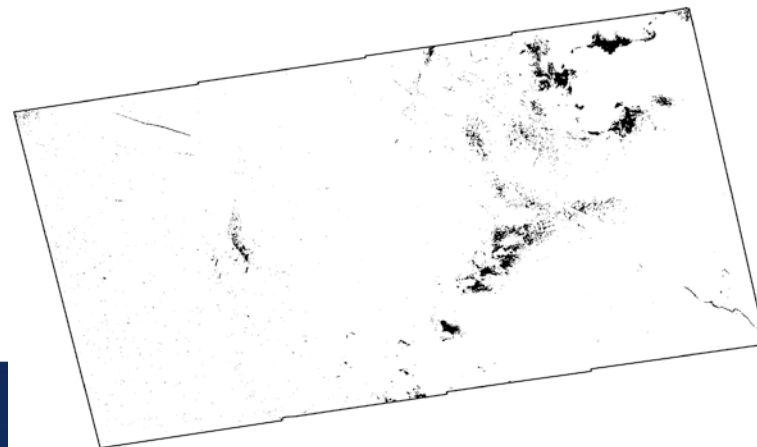
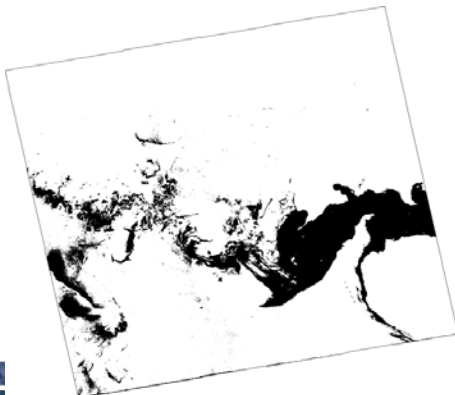




21 May, 2010 Envisat ASAR (C)



13 Sept., 2011 Cosmo-SkyMed SAR (X)

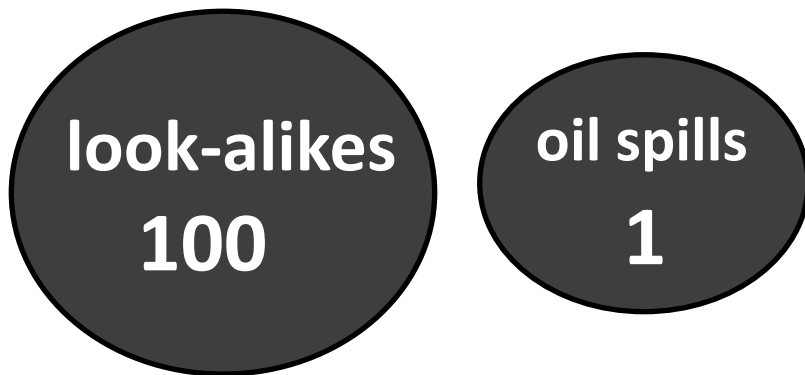


Why imbalance?

- Requirements of automation.

All targets extracted by segmentation algorithm must be processed by program.

1.6 million targets are extracted from 336 SAR images



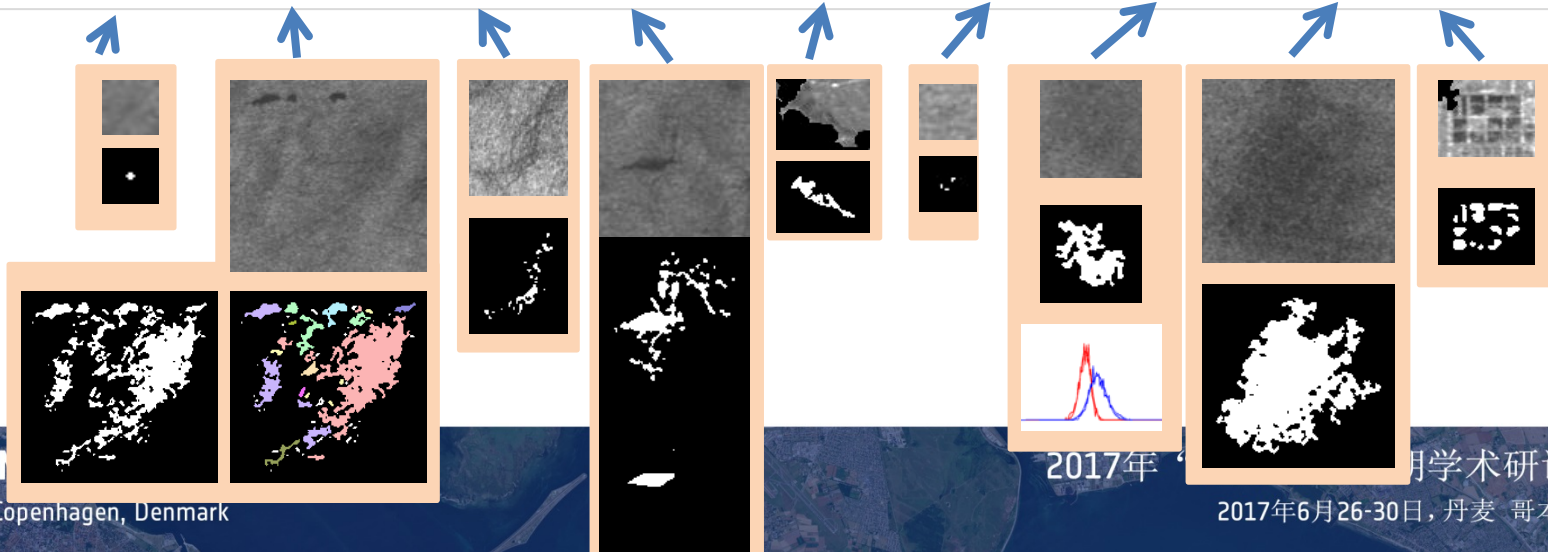
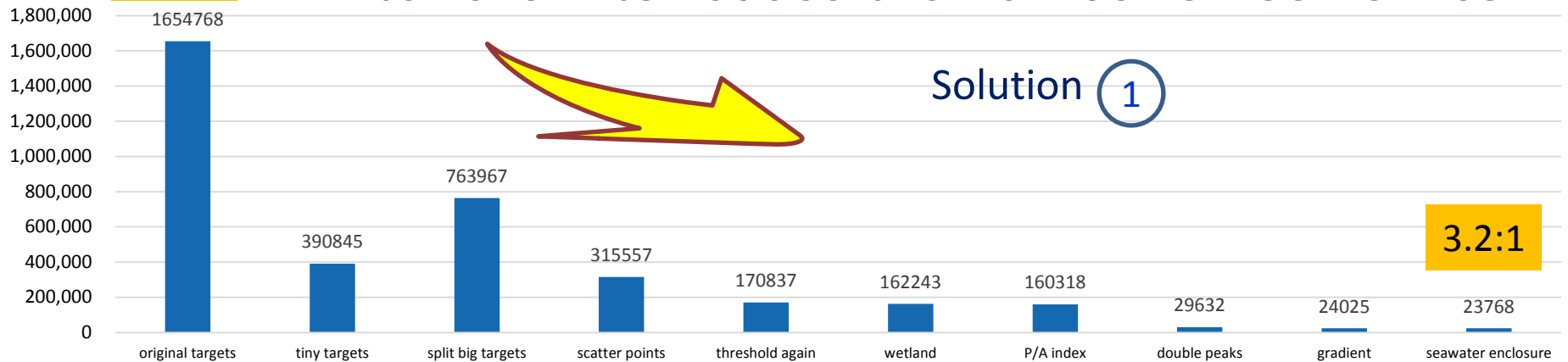
Satellite SAR	Number	Geo Correction	Product Type
Envisat / ASAR	138	No	IMP, WSM
ERS-1、2 / SAR	63	No	IMP/PRI
Cosmo Sky-MED / SAR	135	Corrected (Mercator Proj.)	WR, HI
Total	336		

100:1

Filter chain to reduce the number of look-alikes

Solution ①

3.2:1



2017 DRAGON

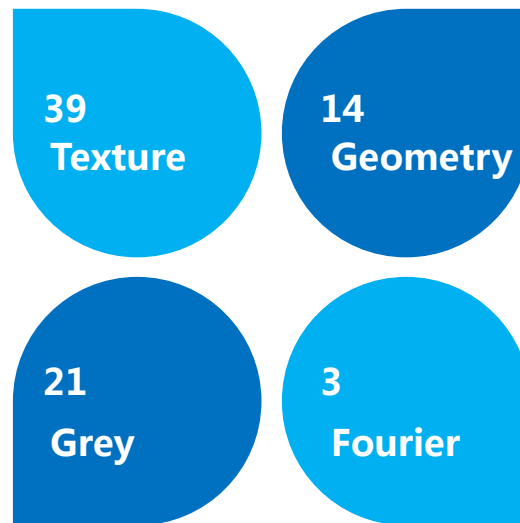
26-30 June 2017 | Copenhagen, Denmark

2017年“海洋学术研讨会”

2017年6月26-30日, 丹麦 哥本哈根

Solution ②

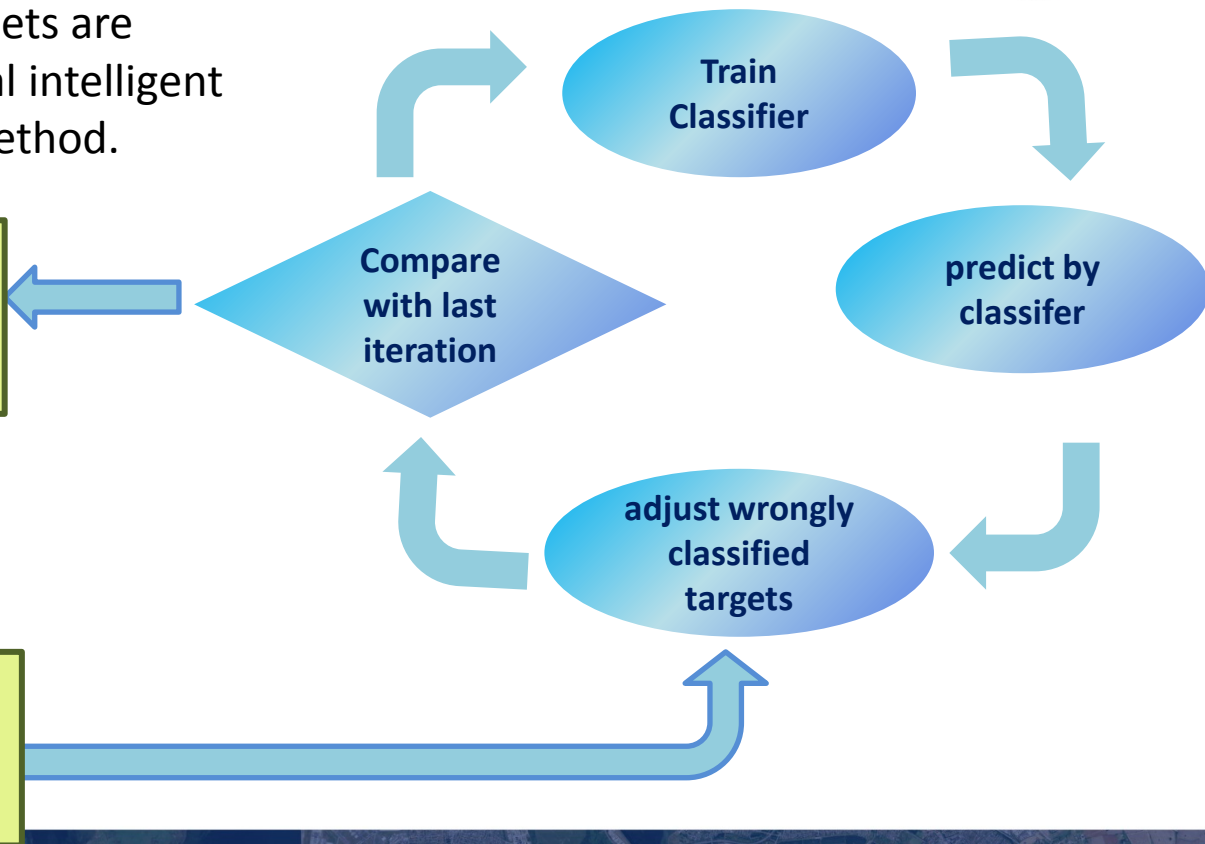
- 77 features are calculated for each dark target.
- The features are collected from the literatures of oil spill detection and image processing.

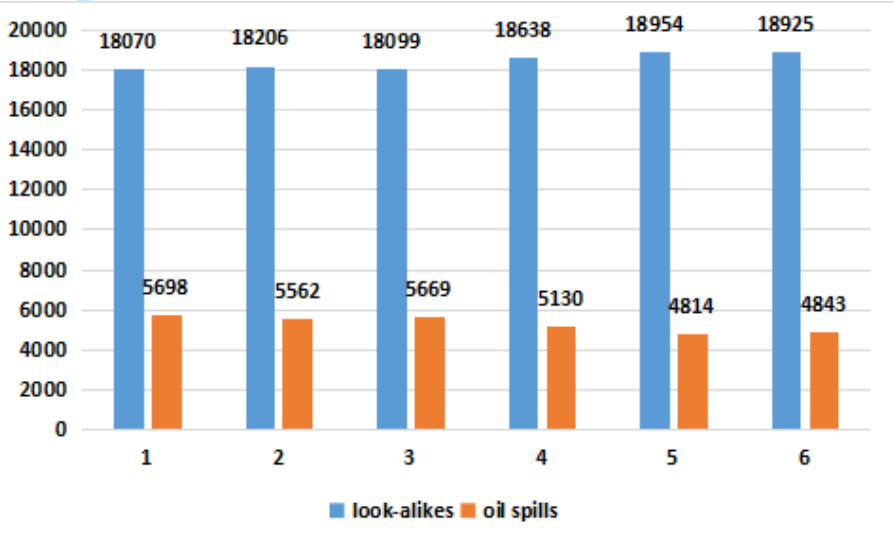


The class labels of targets are assigned by an artificial intelligent iteratively training method.

targets with class label assigned
3.9:1

targets roughly classified by experts.
3.2:1





		classes assigned by experts	
		oil ($Y=Y_y+Y_n$)	look-alike ($N=N_y+N_n$)
predict by classify	oil ($y=Y_y+N_y$)	Y_y	N_y
	look-alikes ($n=Y_n+N_n$)	Y_n	N_n

Recognition Rate

$$\text{识别率} = (Y_y + N_n) / (Y_y + N_y + Y_n + N_n)$$

Detection Rate

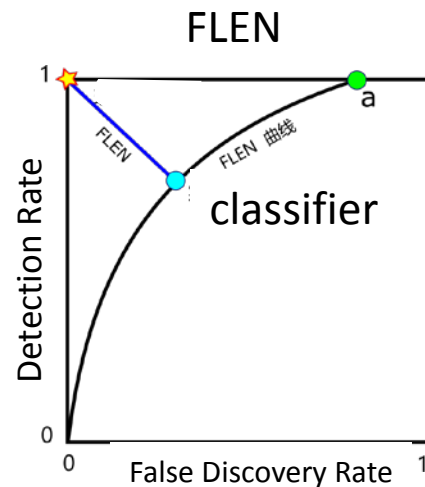
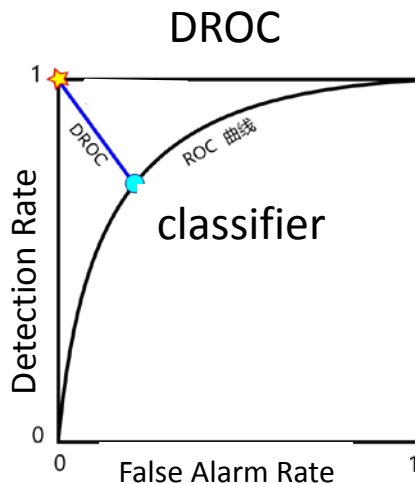
$$\text{检测率} = Y_y / (Y_y + Y_n)$$

False Alarm Rate (FAR)

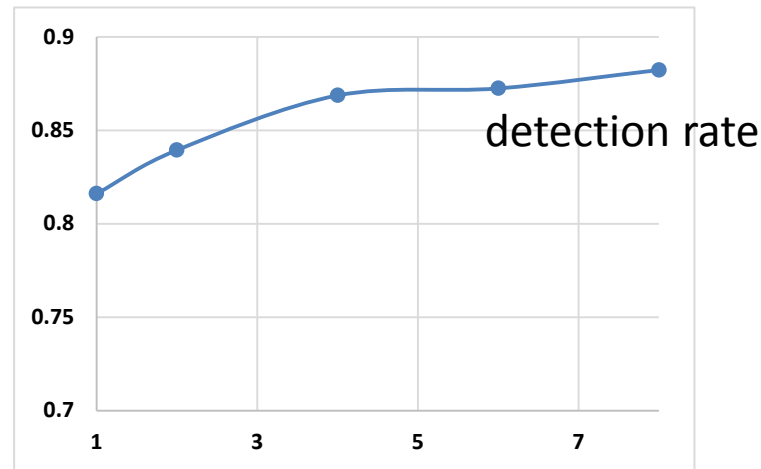
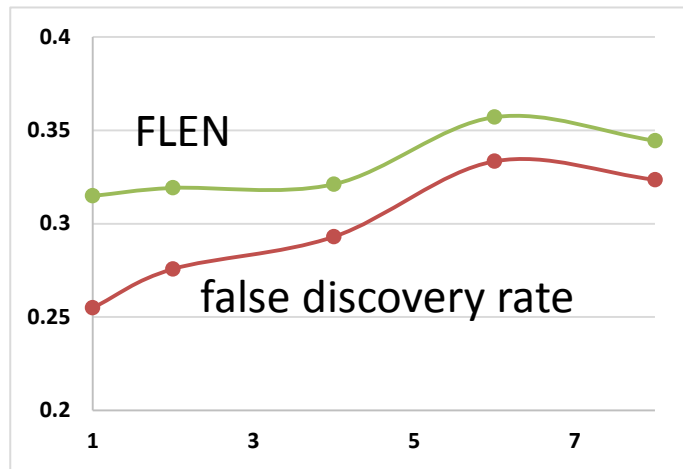
$$\text{虚警率} = N_y / (N_y + N_n)$$

False Discovery Rate (FDR)

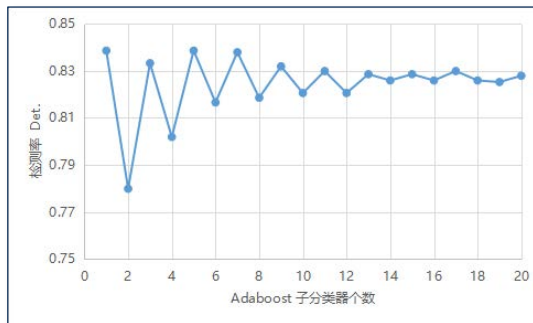
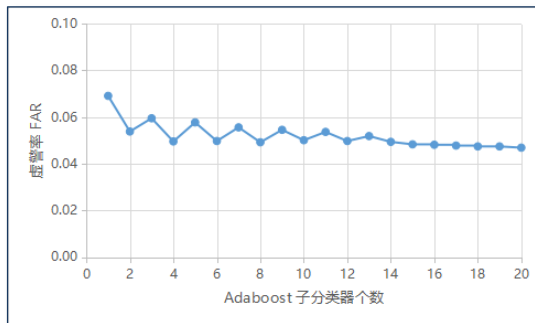
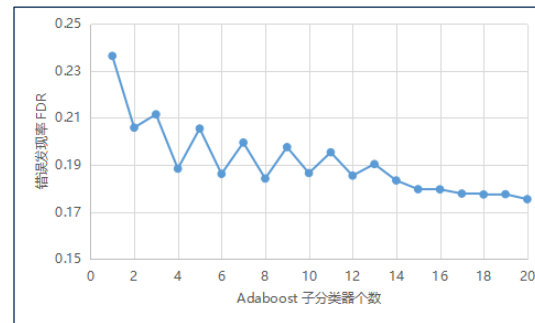
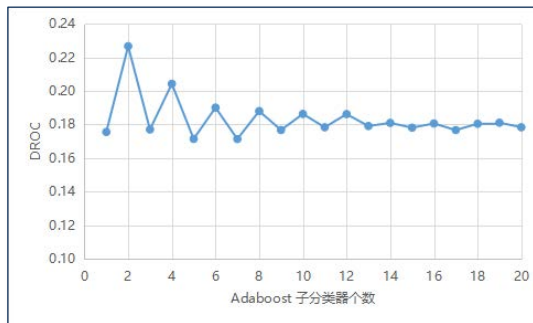
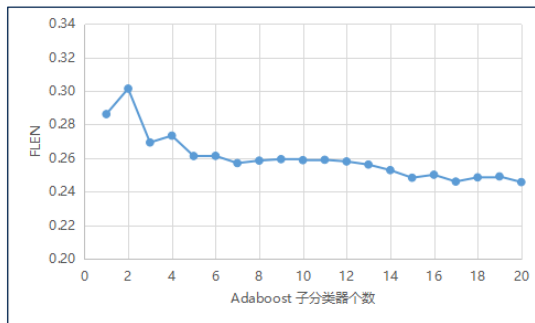
$$\text{错检率} = N_y / (Y_y + N_y)$$



oversample the oil spill targets



AdaBoost MLP 77:70:15:2



Area weighted AdaBoost MLP (MLP-AA) 77:70:15:2

$$w_{F,i} = Area_i^{1/4}$$

	det. Rt (%)	False alarm (%)	False Discovery (%)	Accuracy (%)	FLEN
avg & stddev	80.8±0.5	5.0±0.1	19.8±0.4	92.1±0.1	0.276±0.003
best FLEN	81.2	5.0	19.5	92.3	0.271

Summary

- High classification performance is achieved in a fully automatic SAR oil spill detection system. The techniques used in the system are explained in the point view of imbalance problems.
- Adaptive threshold based on multi-scale background normalization and its pre-processing ensures the extracted dark areas cover all oil spills. The filter chain with simple rules reduces the look-alikes dramatically and therefore decrease the degree of imbalance between oil spills and look-alikes.
- 77 features are used for classification. They are most comprehensive feature set so far.
- Class labels are assigned by an artificial intelligent iteritivy training method, which makes the targets more easy to sepearate therefore reduce the complexity of classification problem.
- For 2-hidden layer MLP. The oversample is not helpful to improve the classification pefermance..
- Adaboost two-hidden-layer MLP can significantly reduce false discover rate (FDR) while keeping DR in high level.

Thank you