Adaptive Remote Sensing Image Attribute Learning for Active Object Detection

Nuo Xu

NLPR, Institute of Automation, Chinese Academy of Sciences

University of Chinese Academy of Sciences

INTRODUCTION

Traditional object recognition approaches are limited due to the passive nature:

- In imaging procedure, images are acquired driven by visual inspection rather than object recognition performance.
- In object recognition procedure, images are directly used for training or testing without active adjustment.

MOTIVATION

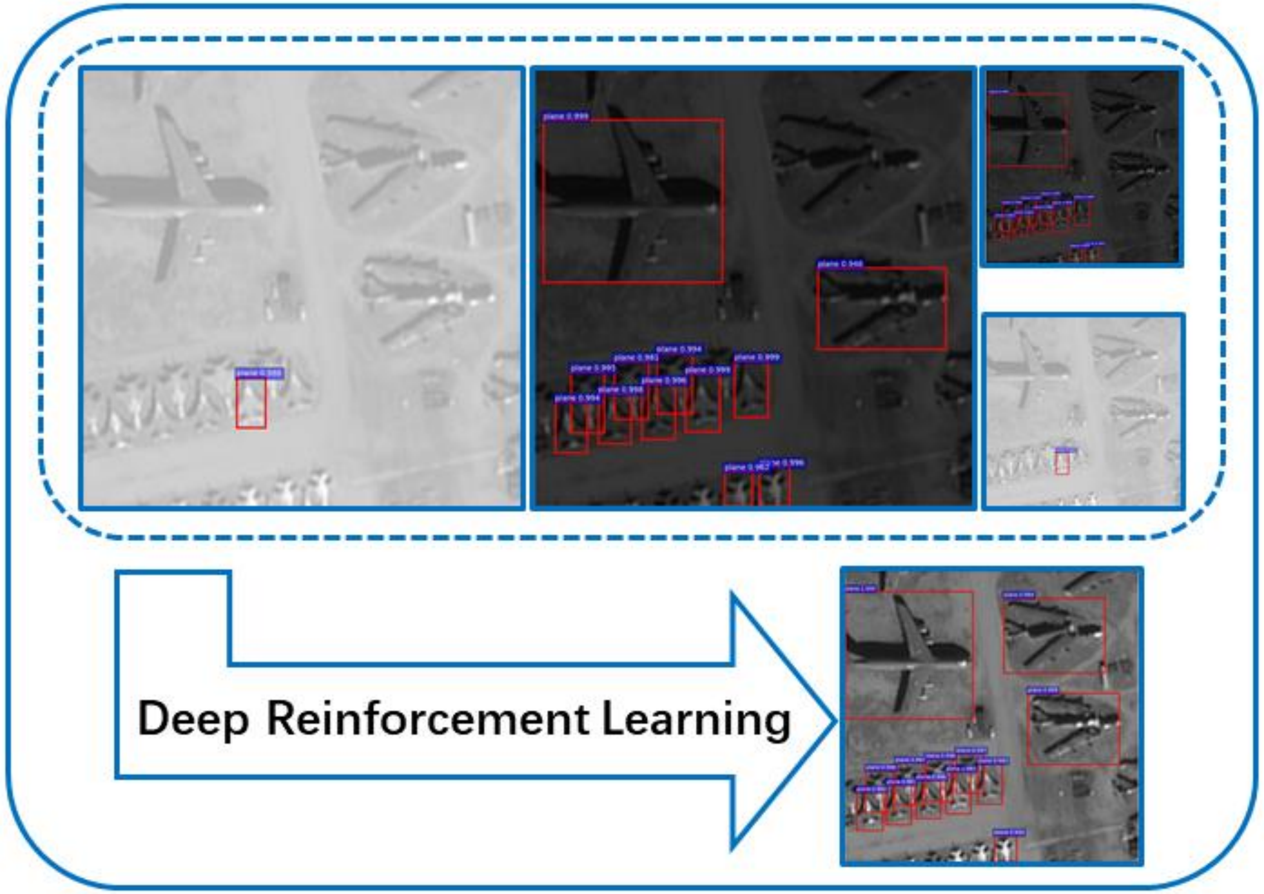

Fig. 1 Motivation. Due to the limitation of imaging configuration and environmental changes, the detection performance of low-quality images is not good. Therefore, it is necessary to adaptively learn image attributes to improve detection performance.

METHODOLOGY

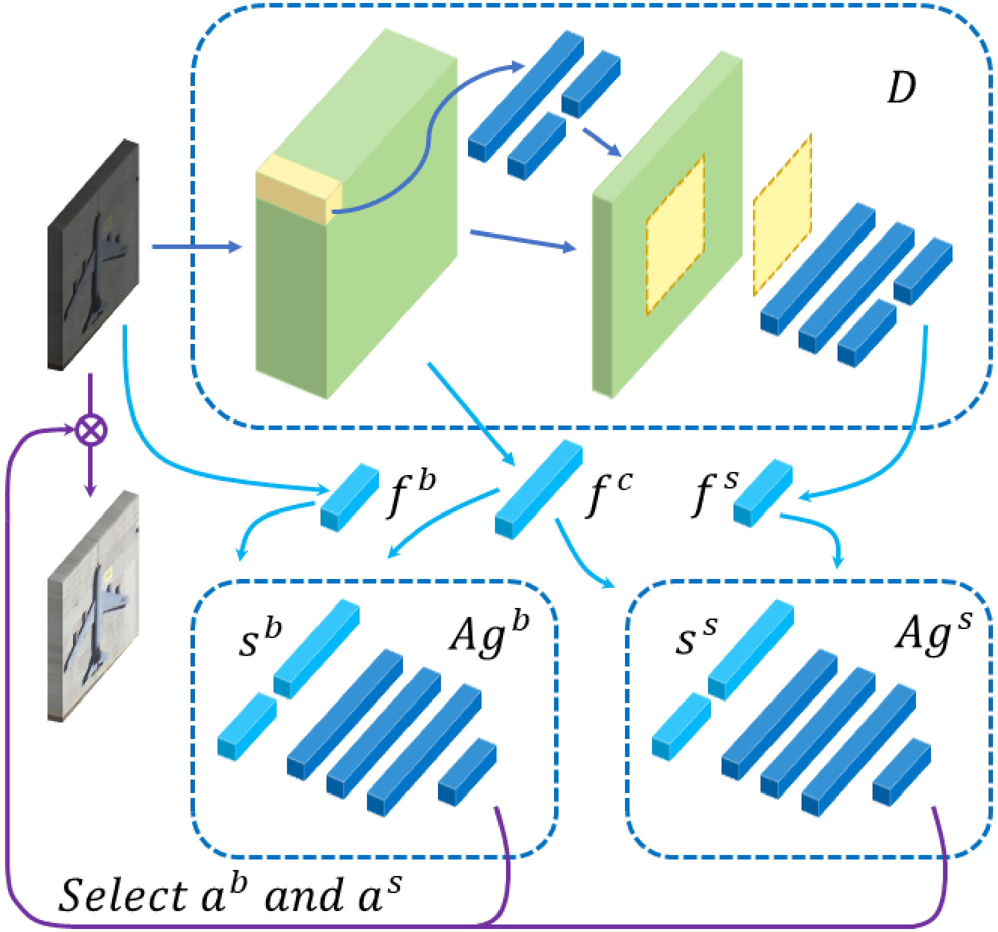

Fig. 2 Overview of RL-AOD.

Environment *D*

Detector Faster RCNN.

Agent Ag^b, Ag^s

Fully connected network.

State $s^b(t)$, $s^s(t)$

Features of the image.

$s^b(t)=\{f^c, f^b\}$

$s^s(t)=\{f^c, f^s\}$

Reward $r^b(t)$, $r^s(t)$

+1, if performance is improved; otherwise, -1.

Action $a^b(t)$, $a^s(t)$

$a^b(t)$=**{brighten, darken}**

$a^s(t)$=**{zoom in, zoom out}**

METHODOLOGY

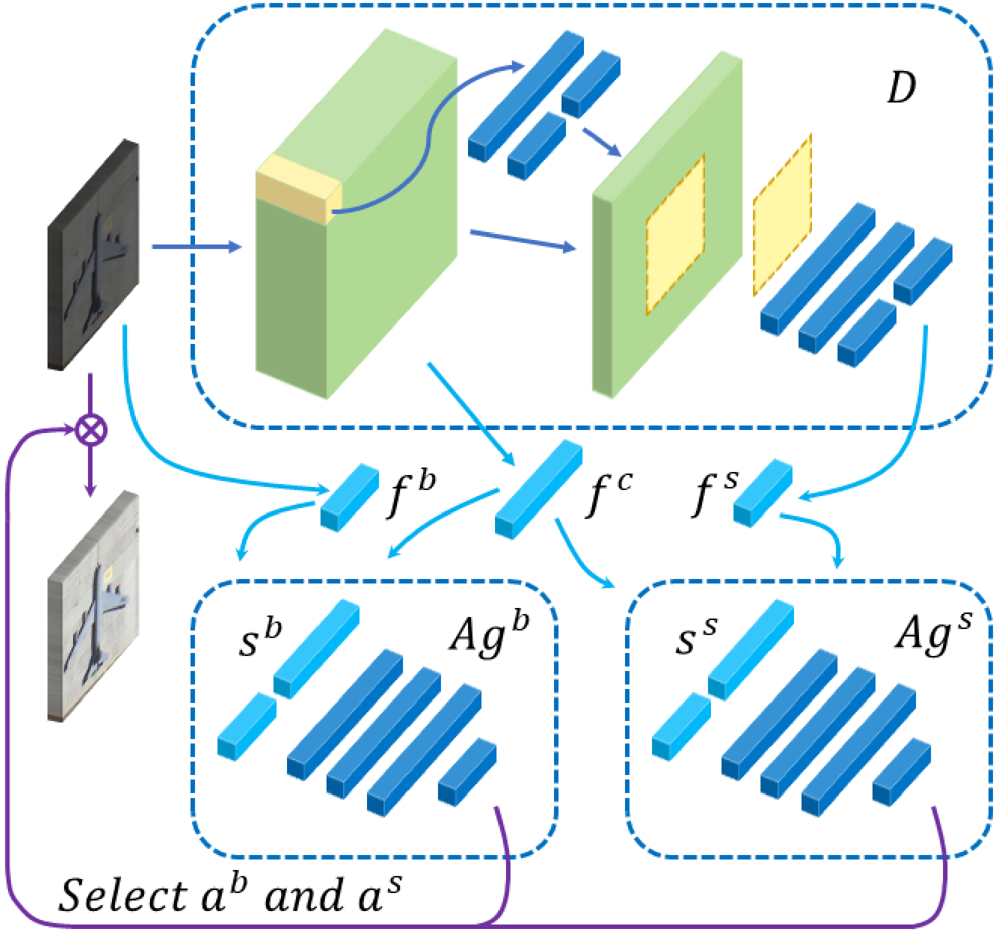

Fig. 2 Overview of the proposed approach.

Scale feature f^s

f^s is statistical histogram of the objects' area, of 64 dimensions.

Brightness feature f^b

f^b is statistical histogram of the component V in HSV color space, of 64 dimensions.

Contextual feature f^c

f^c is the middle output of Faster RCNN (feature maps before RoI Pooling layer). By taking the average along channel, f^c is straightened into a 512-dimension vector.

METHODOLOGY

- Adaptive image attribute learning is described as follows:
- At step t, an image $I(t)$ is represented by contextual feature f^c, brightness feature f^b and scale feature f^s.
- Agent selects action $a^b(\mathrm{t})$ and $a^s(\mathrm{t})$ based on state $s^b(t)$ and $s^s(t)$, respectively. Image $I'(t+1)$ is obtained by applying action $a^b(\mathrm{t})$ on $I(t)$. Image $I(t+1)$ is obtained by applying action $a^s(\mathrm{t})$ on $I'(t+1)$.
- Reward $r^b(\mathrm{t}+1)$ is calculated based on the change of detection performance between $I'(t+1)$ and $I(t)$. Reward $r^s(\mathrm{t}+1)$ is calculated based on the change of detection performance between $I(t+1)$ and $I'(t+1)$.
- $I(T)$ can be obtained, after applying a series of actions to $I(0)$.

EXPERIMENT

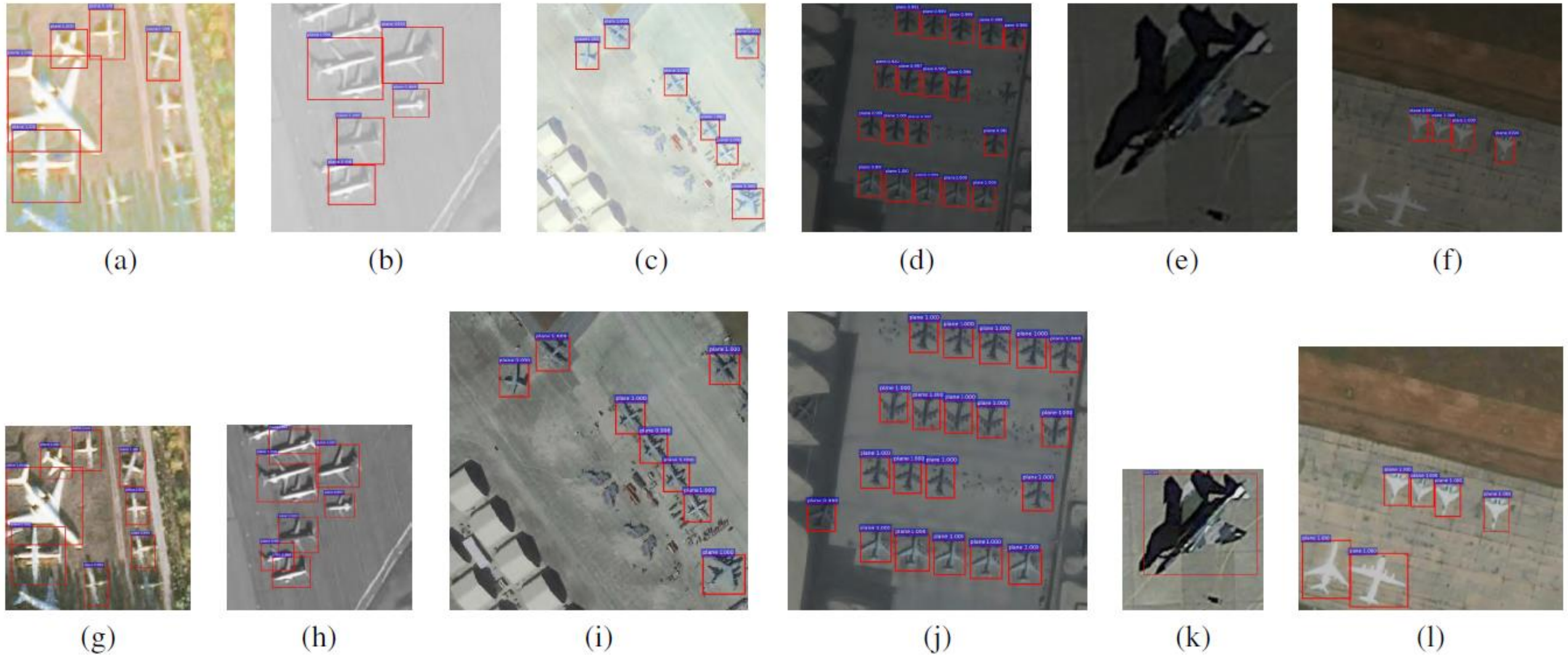

Fig. 3 Detection performance comparison. The images (a, b, c, d, e, f) in the first row are the results on the degraded images. The images (g, h, i, j, k, l) in the second row are the results after adaptive learning by RL-AOD.

EXPERIMENT

Tab. 1 Performance comparison of different methods.

Method+Backbone	AP	AP^{50}	AP^{75}	AP^{S}	AP^{M}	AP^{L}
DPMv5 (benchmark)	-	0.338	-	-	-	-
RetinaNet+VGG16	0.376	0.585	0.431	0.249	0.492	0.526
RetinaNet+ResNet50	0.446	0.674	0.515	0.297	0.591	0.588
RetinaNet+ResNet101	0.503	0.732	0.596	0.338	0.654	0.681
SSD321+ResNet101	0.417	0.661	0.475	0.200	0.594	0.703
DSSD321+ResNet101	0.426	0.666	0.485	0.196	0.610	0.739
YOLOv2+DarkNet19	0.407	0.632	0.472	0.202	0.573	0.701
YOLOv3+DarkNet53	0.491	0.808	0.553	**0.441**	0.574	0.401
R-FCN+ResNet50	0.422	0.705	0.461	0.223	0.576	0.690
R-FCN+ResNet101	0.427	0.713	0.464	0.225	0.582	0.694
Faster RCNN+VGG16	0.441	0.750	0.469	0.273	0.587	0.652
Faster RCNN+Res50	0.455	0.768	0.482	0.273	0.601	0.700
Faster RCNN+Res101	0.479	0.784	0.525	0.300	0.626	0.703
RL-AOD+VGG16	0.530	**0.822**	0.608	0.355	**0.674**	**0.751**
RL-AOD+ResNet50	0.519	0.815	0.590	0.346	0.661	0.734
RL-AOD+ResNet101	**0.531**	**0.822**	**0.612**	0.361	0.664	0.750

Thanks