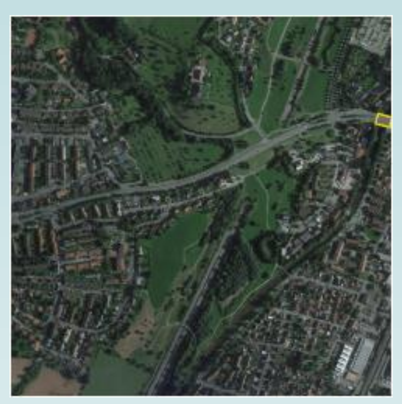
Cascade Saliency Attention Network for Object Detection in Remote Sensing Images

Dayang Yu, Rong Zhang, Shan Qin*
University of Science and Technology of China, Hefei, China

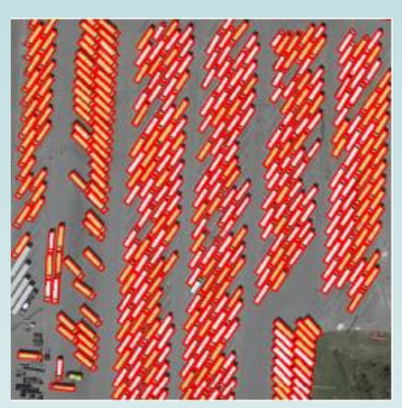
Problems

Challenges of object detection in remote sensing images

- Complex Background. Due to the high resolution of remote sensing images, a large number of complex backgrounds overwhelm the objects.
- Dense Arrangement. In the special bird-view perspective from which remote sensing images are taken, some categories of objects are often densely arranged.
- Large-scale Variations. Objects in remote sensing images show extreme variation in scales, which is common between and within categories.



Complex background



Dense arrangement

Large-scale variations

Motivation

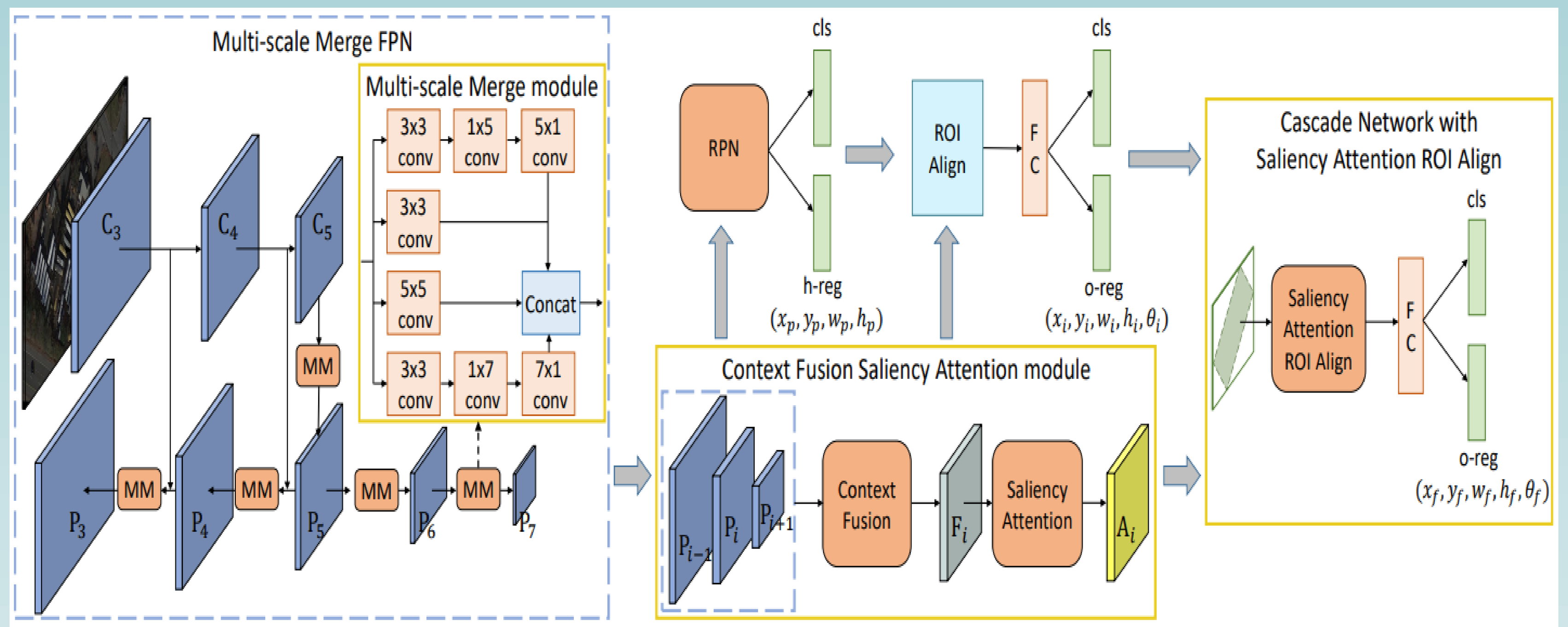
- Merely enhancing object features may cause background features to be enhanced at the same time.
- Rotate ROI Align loses angle information of objects and ROI Align may contains multiple objects.
- Traditional single-scale convolutional layers cannot effectively capture features at different scales.

Contribution

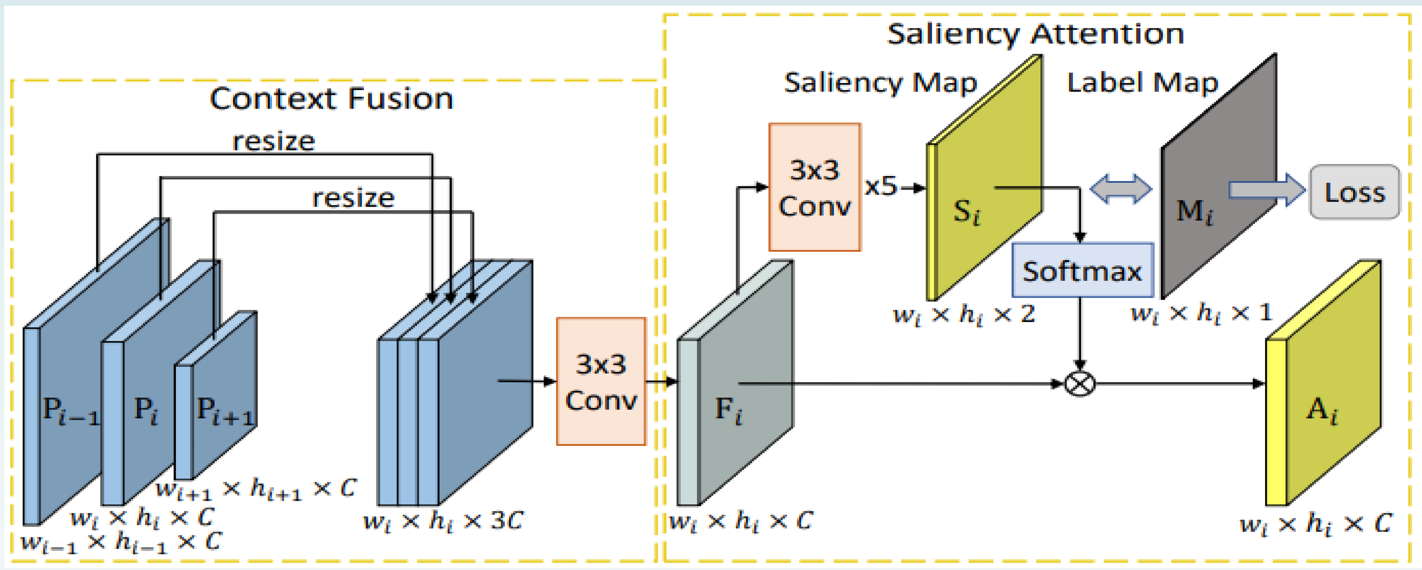
- We propose a Context Fusion Saliency Attention module that combines context and pixel attention to mitigate the adverse impact of complex backgrounds
- We develop a Saliency Attention ROI Align for increasing saliency of the central object on ROI. It effectively reduces mutual interference from densely arranged objects.
- We devise a Multi-scale Merge module to handle largescale variations by learning multi-scale representations.

Methodology

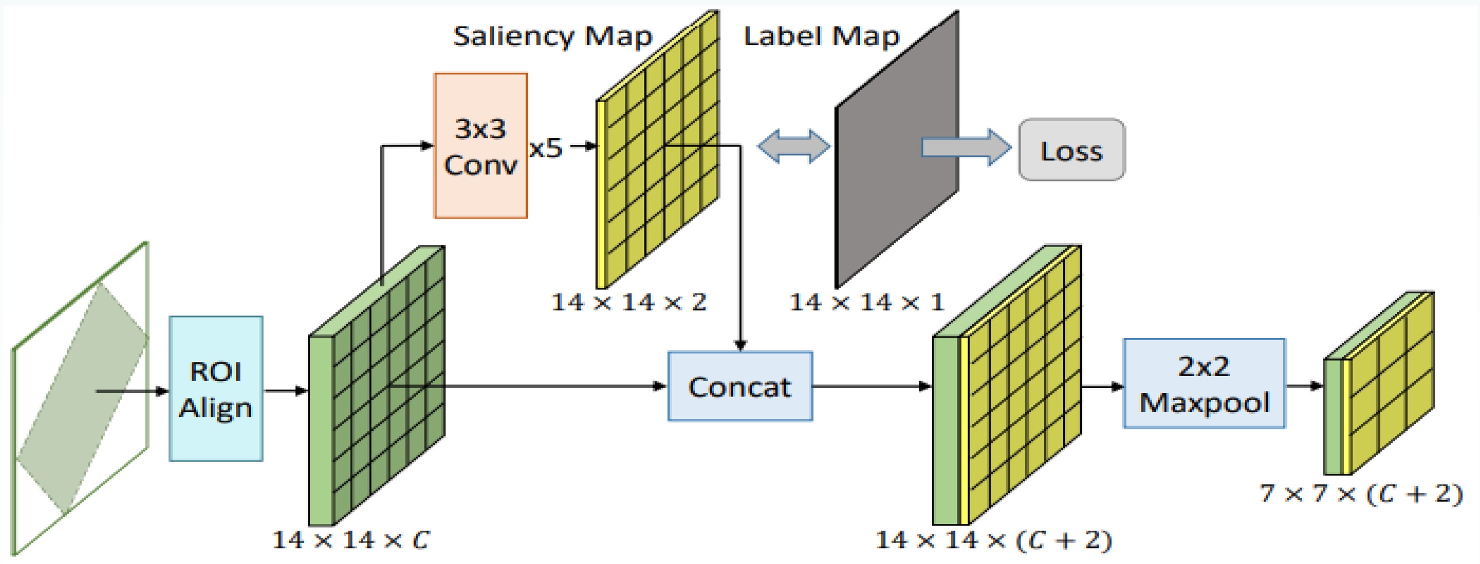
The framework of proposed network. The network contains three main subparts: Multi-scale Merge module in FPN, Context Fusion Saliency Attention module, and cascade network with Saliency Attention ROI Align.



The architecture of Context Fusion Saliency Attention module. This module first fuses multi-layer features and then performs pixel attention on the context fusion feature map.



The architecture of cascade network with Saliency Attention ROI Align. At cascade stage, the instance segmentation is used on ROIs extracted by ROI Align.



Experiments

Results on DOTA

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
FR-H [17]*	47.16	61.00	9.80	51.74	14.87	12.80	6.88	56.26	59.97	57.32	47.83	48.70	8.23	37.25	23.05	36.29
FR-O [13]*	79.42	77.13	17.70	64.05	35.30	38.02	37.16	89.41	69.64	59.28	50.30	52.91	47.89	47.40	46.30	52.93
R^2CNN [27]	80.94	65.67	35.34	67.44	59.92	50.91	55.81	90.67	66.92	72.39	55.06	52.23	55.14	53.35	48.22	60.67
ICN [12]*	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
ROI Trans [11]*	88.64	78.52	43.44	**75.92**	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
CAD-Net [9]*	87.80	82.40	49.40	73.50	71.10	63.50	76.60	**90.90**	79.20	73.30	48.40	60.90	62.00	67.00	**62.20**	69.90
R^2CNN++ [10]	89.66	81.22	45.50	75.10	68.27	60.17	66.83	**90.90**	80.69	86.15	64.05	63.48	65.34	68.01	62.05	71.16
R3Det [28]	89.24	80.81	51.11	65.62	70.67	76.03	78.32	90.83	**84.89**	84.42	65.10	57.18	68.10	68.98	60.88	72.81
Zhu et al. [22]	89.67	76.77	51.28	71.65	73.11	77.18	79.54	90.79	79.01	84.54	**66.51**	64.71	73.97	67.73	58.40	73.66
Ours	**89.86**	**83.52**	**51.98**	71.58	**75.68**	**80.34**	**87.26**	90.50	80.92	**86.61**	66.11	**67.44**	**74.53**	**69.55**	57.84	**75.58**

Results of ablation study

Baseline	MM	CFSA	SA-ROI Align cas	mAP
√				72.23
√	√			73.13
√		√		73.21
√			√	74.69
√	√	√	√	75.58

Results on HRSC2016

Method	mAP
BL2 [29]	69.60
R^2CNN [27]	73.07
RC1 & RC2 [29]	75.70
R2PN [30]	79.60
RRD [31]	84.30
ROI Trans [11]	86.20
R3Det [28]	89.33
Ours	**93.43**

Qualitative results