



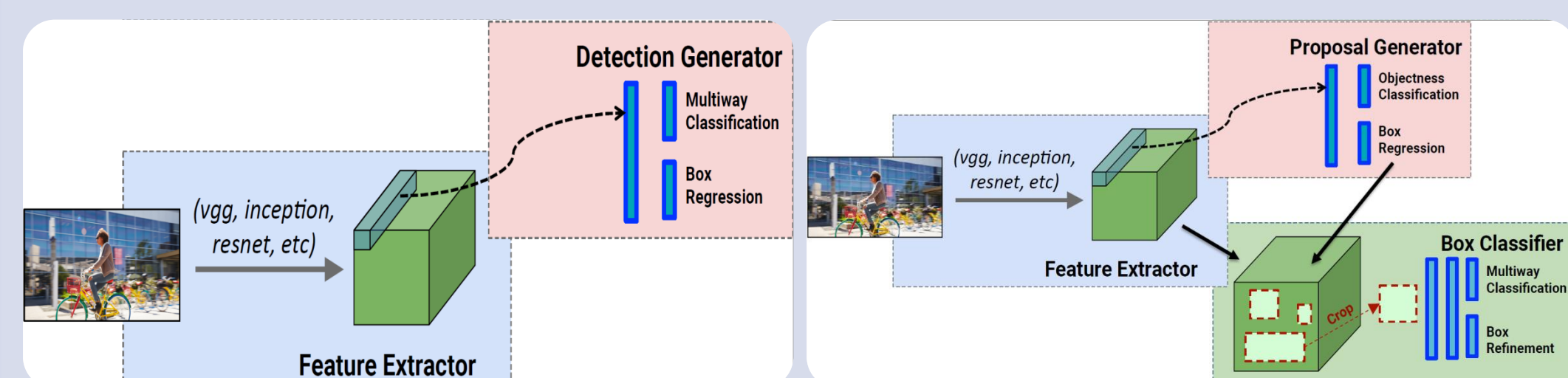
Single-Shot Refinement Neural Network for Object Detection

Shifeng Zhang¹, Longyin Wen², Xiao Bian², Zhen Lei¹, Stan Z. Li¹

¹CBSR & NLPR, Institute of Automation, Chinese Academy of Sciences ²GE Global Research



Motivation



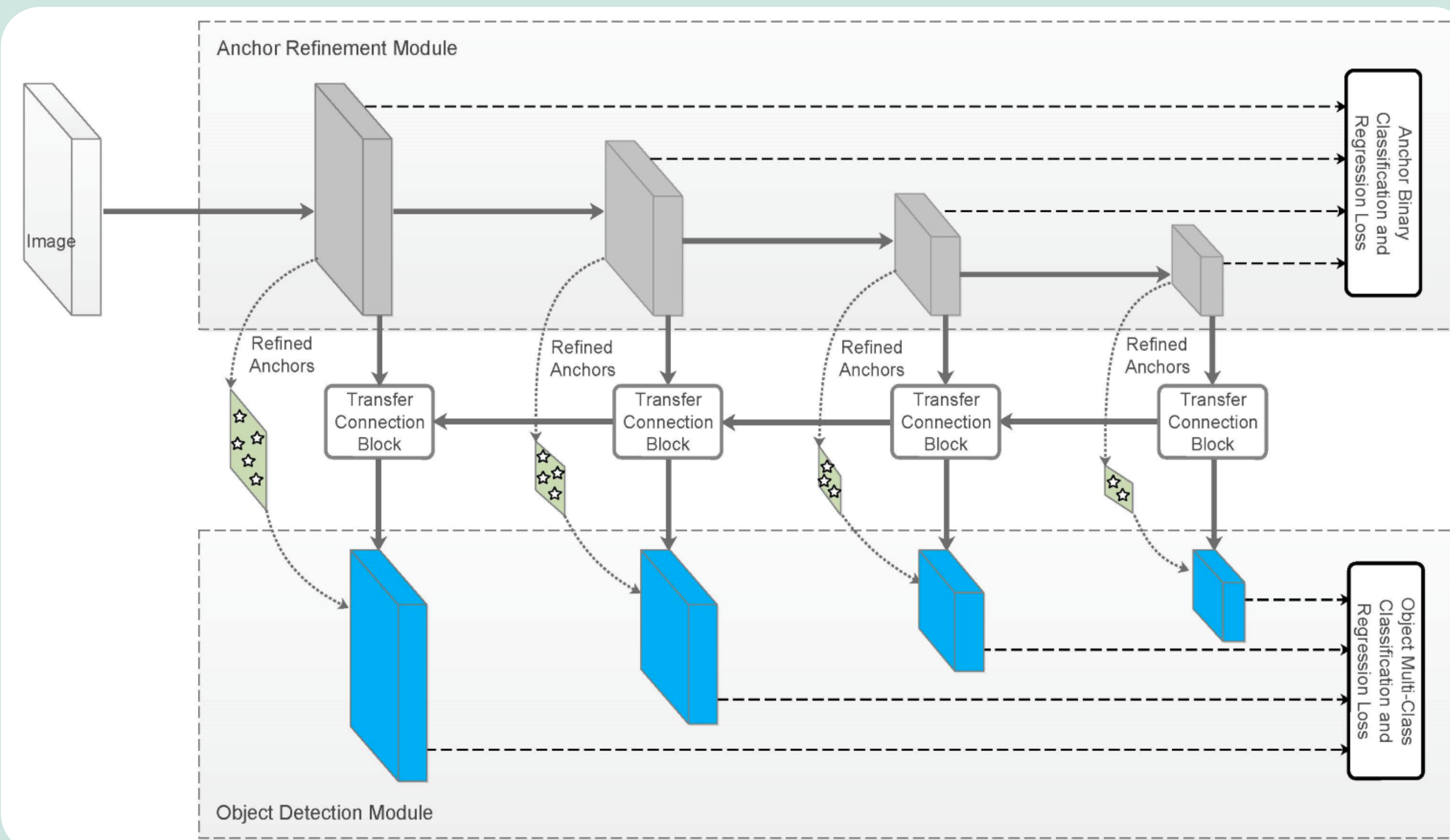
(a) One-Stage: high efficiency (b) Two-Stage: high accuracy

- The one-stage approach has the advantage of high efficiency
- While its detection accuracy is usually behind that of the two-stage approach
- Since two-stage methods have 3 advantages over the one-stage methods:
 - *two-stage structure to handle class imbalance*
 - *two-step regression*
 - *two-stage features to describe the objects*

Contribution

- Introduce a new one-stage framework for object detection, composed of two inter-connected modules, i.e., the ARM and the ODM.
- Design the TCB to transfer the features in the ARM to handle detection task in the ODM.
- Achieve better accuracy than two-stage methods and maintain comparable efficiency of one-stage.

Architecture



- **Anchor Refinement Module (ARM)**
 - Filter out negative anchors to reduce search space
 - Adjust the pre-set anchors to provide better initialization
- **Transfer Connection Block (TCB)**
 - Transfer the features in the ARM to complete more accurate regression and classification in the ODM
 - Integrate large-scale context by adding the high-level features to the transferred features using the deconvolution operation
- **Object Detection Module (ODM)**
 - Take the refined anchors as input to further improve the regression accuracy and predict multi-class label

Model Analysis

Component	RefineDet320			
negative anchor filtering?	✓			
two-step cascaded regression?	✓	✓		
transfer connection block?	✓	✓	✓	
mAP (%) on VOC07 test	80.0	79.5	77.3	76.2

Accuracy vs. Speed

Method	VOC07 test mAP	FPS (Titan X)	Number of Boxes	Input Resolution
Faster R-CNN (VGG16)	73.2	7	~6000	~1000 x 600
YOLO (GoogLeNet)	63.4	45	98	448 x 448
YOLOv2 (Darknet19)	78.6	40	1445	544 x 544
SSD300* (VGG16)	77.2	46	8732	300 x 300
SSD512* (VGG16)	79.8	19	24564	512 x 512
RefineDet320 (VGG16)	80.0	40	6375	320 x 320
RefineDet512 (VGG16)	81.8	24	16320	512 x 512

Results on Benchmarks

Method	VGG-16							ResNet-101		
	VOC07 test		VOC12 test		COCO test-dev15			COCO test-dev15		
	0712	0712+ COCO	0712	0712+ COCO	Trainval35k			trainval35k		
	0.5	0.5	0.5	0.5	0.5:0.95	0.5	0.75	0.5:0.95	0.5	0.75
RefineDet320	80.0	84.0	78.1	82.7	29.4	49.2	31.3	32.0	51.4	34.2
RefineDet512	81.8	85.2	80.1	85.0	33.0	54.5	35.5	36.4	57.5	39.5
RefineDet320+	83.1	85.6	82.7	86.0	35.2	56.1	37.7	38.6	59.9	41.7
RefineDet512+	83.8	85.8	83.5	86.8	37.6	58.7	40.8	41.8	62.9	45.7