

SSD

SSD: Single Shot MultiBox Detector

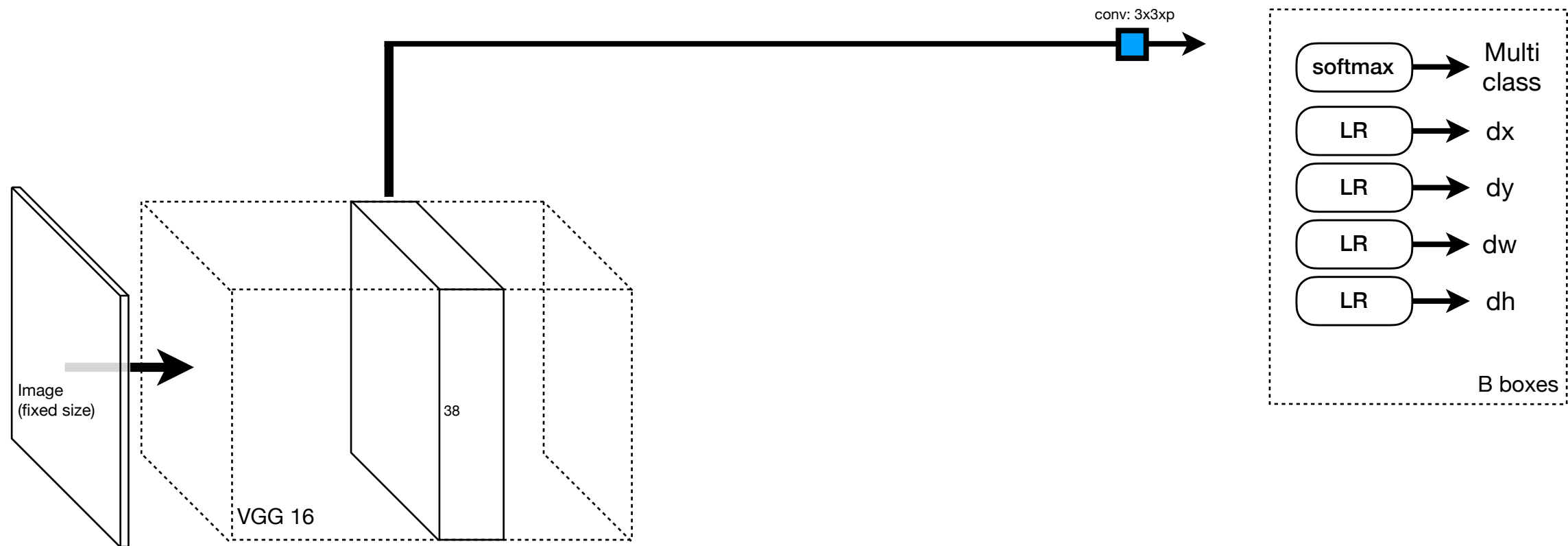
Wei Liu¹, Dragomir Anguelov², Dumitru Erhan³, Christian Szegedy³,
Scott Reed⁴, Cheng-Yang Fu¹, Alexander C. Berg¹

¹UNC Chapel Hill ²Zoox Inc. ³Google Inc. ⁴University of Michigan, Ann-Arbor
¹wliu@cs.unc.edu, ²drago@zoox.com, ³{dumitru,szegedy}@google.com,
⁴reedscot@umich.edu, ¹{cyfu,aberg}@cs.unc.edu

Abstract. We present a method for detecting objects in images using a single deep neural network. Our approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. SSD is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. Experimental results on the PASCAL VOC, COCO, and ILSVRC datasets confirm that SSD has competitive accuracy

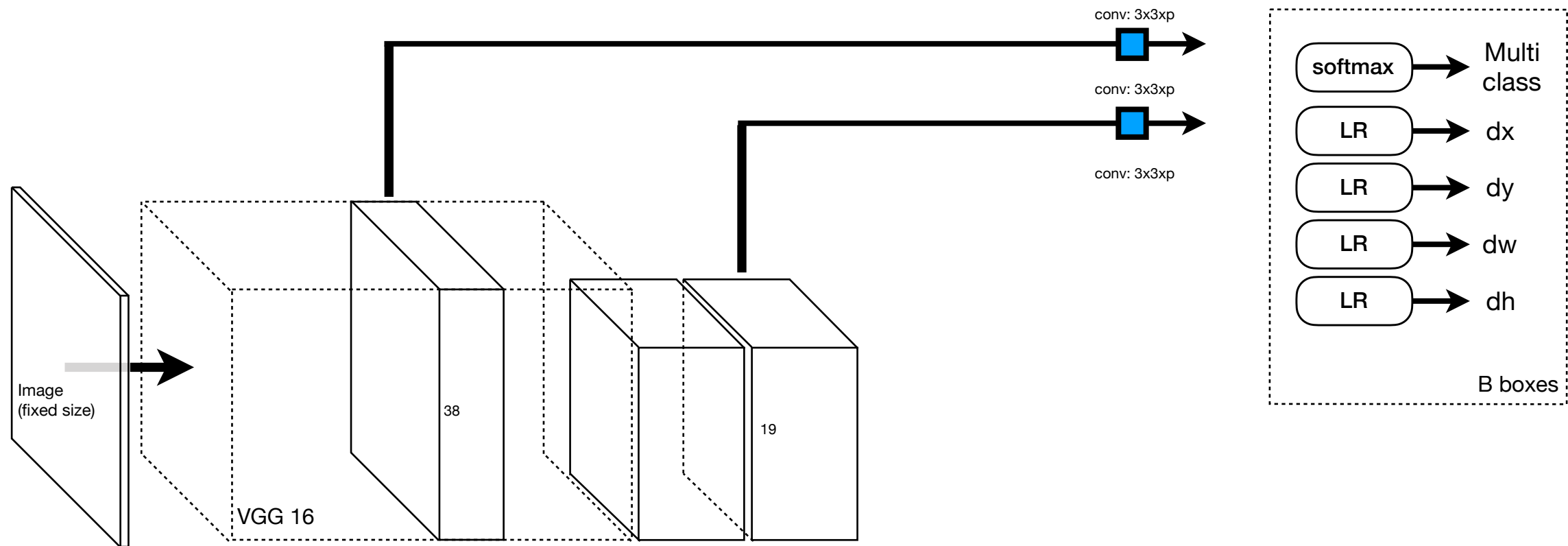
SSD

(Single Shot Multi-box Detection)



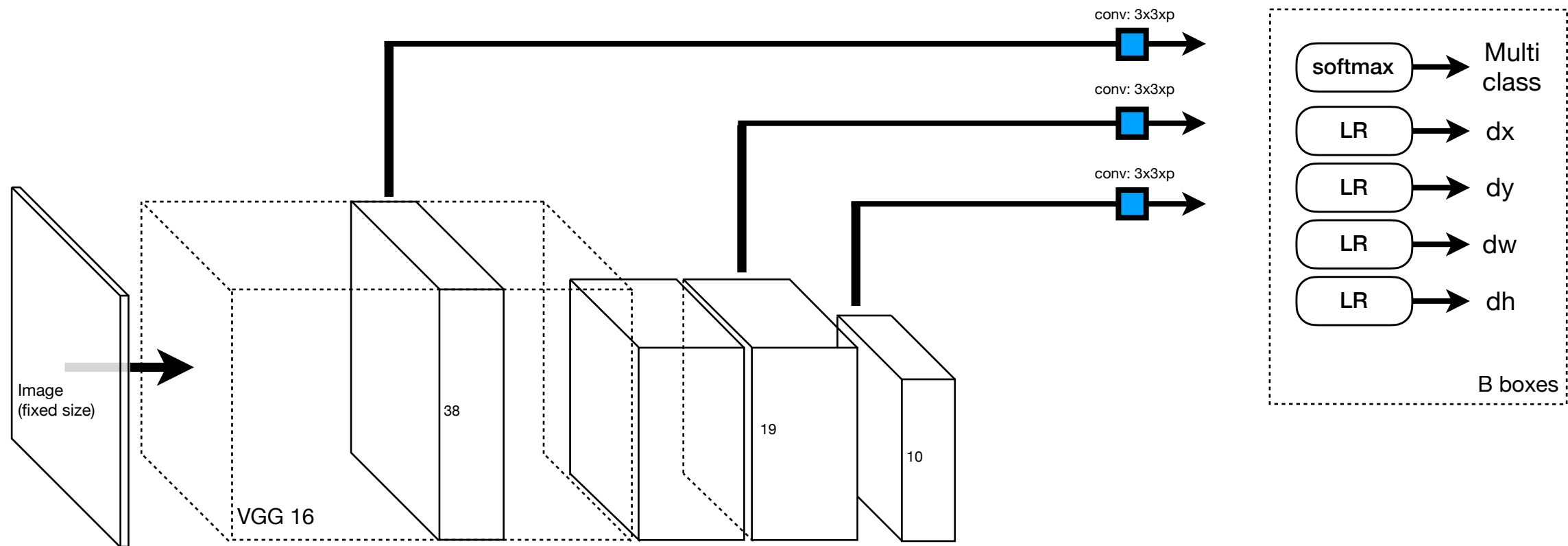
(Single Shot Multi-box Detection)

(Single Shot Multi-box Detection)



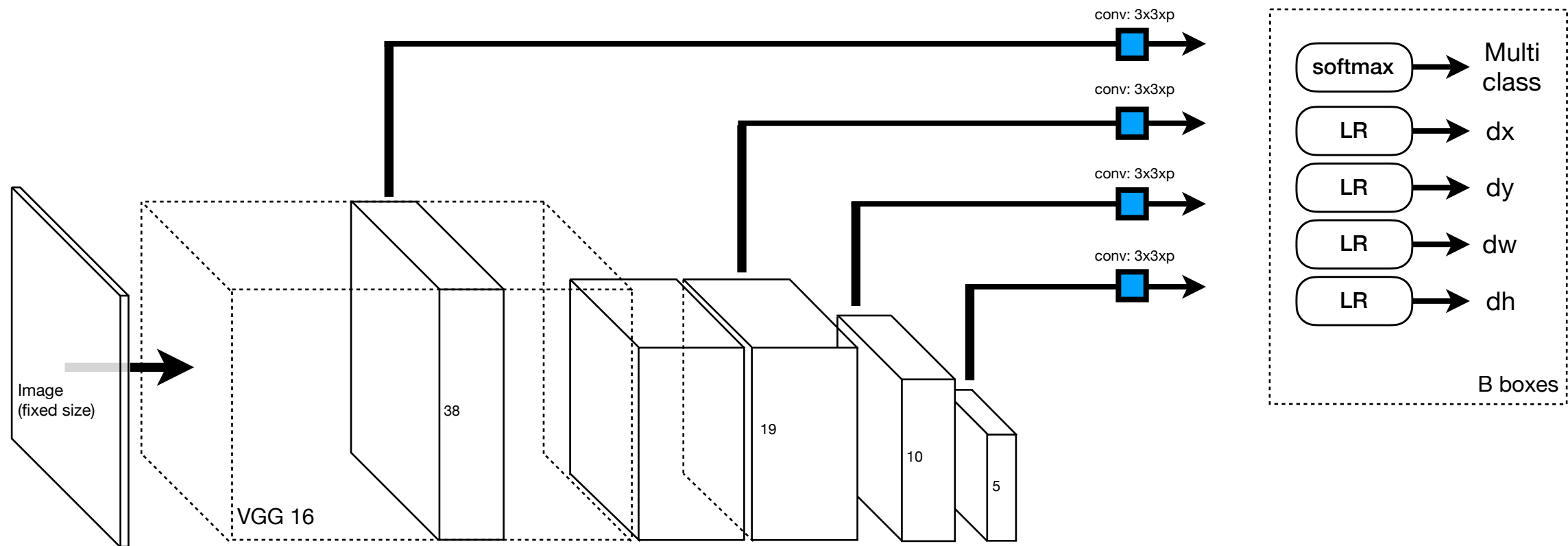
SSD

(Single Shot Multi-box Detection)

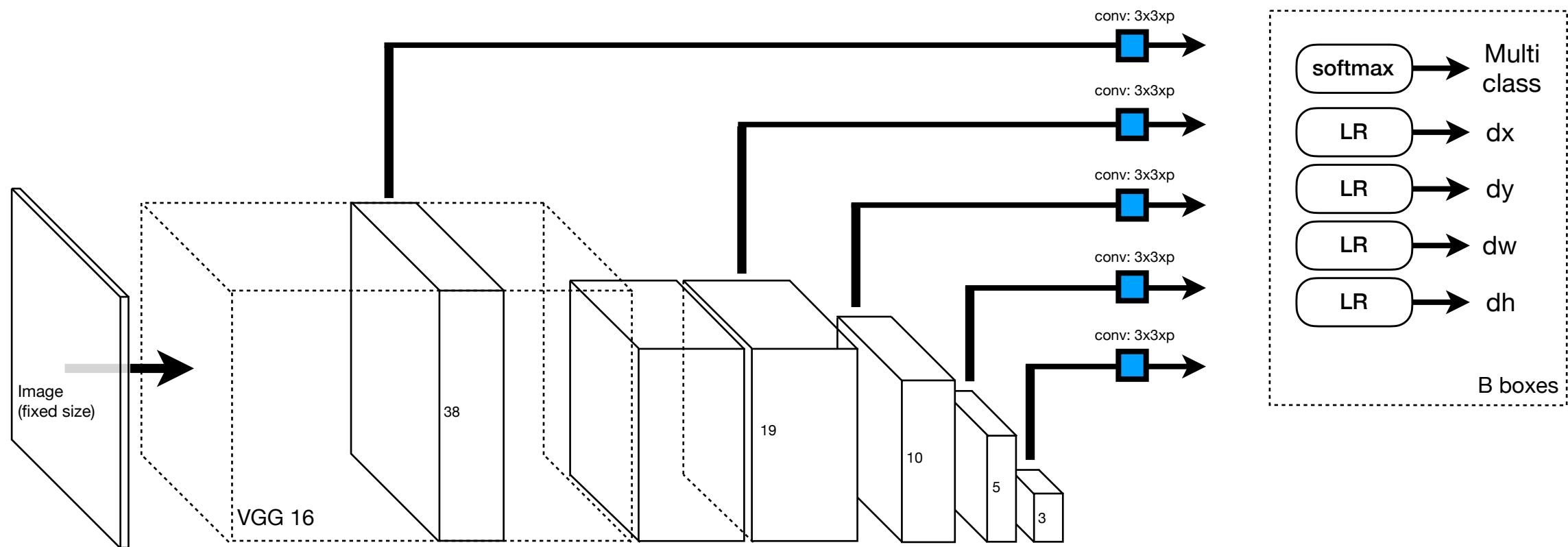


SSD

(Single Shot Multi-box Detection)

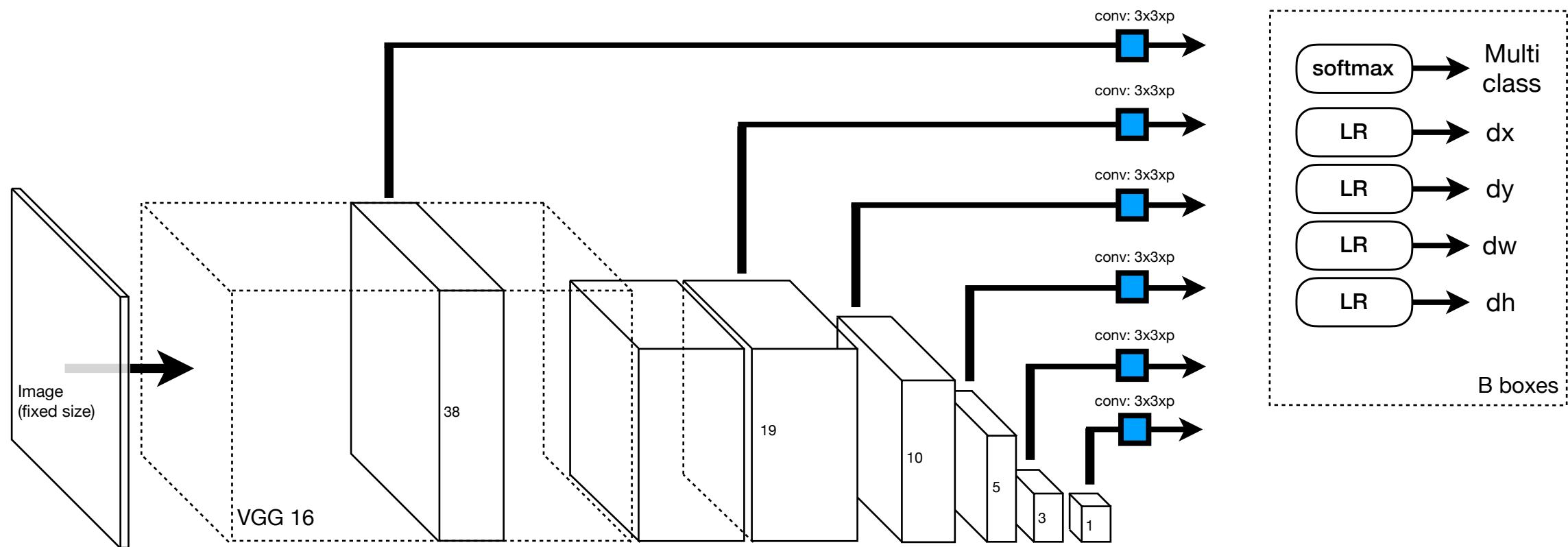


(Single Shot Multi-box Detection)



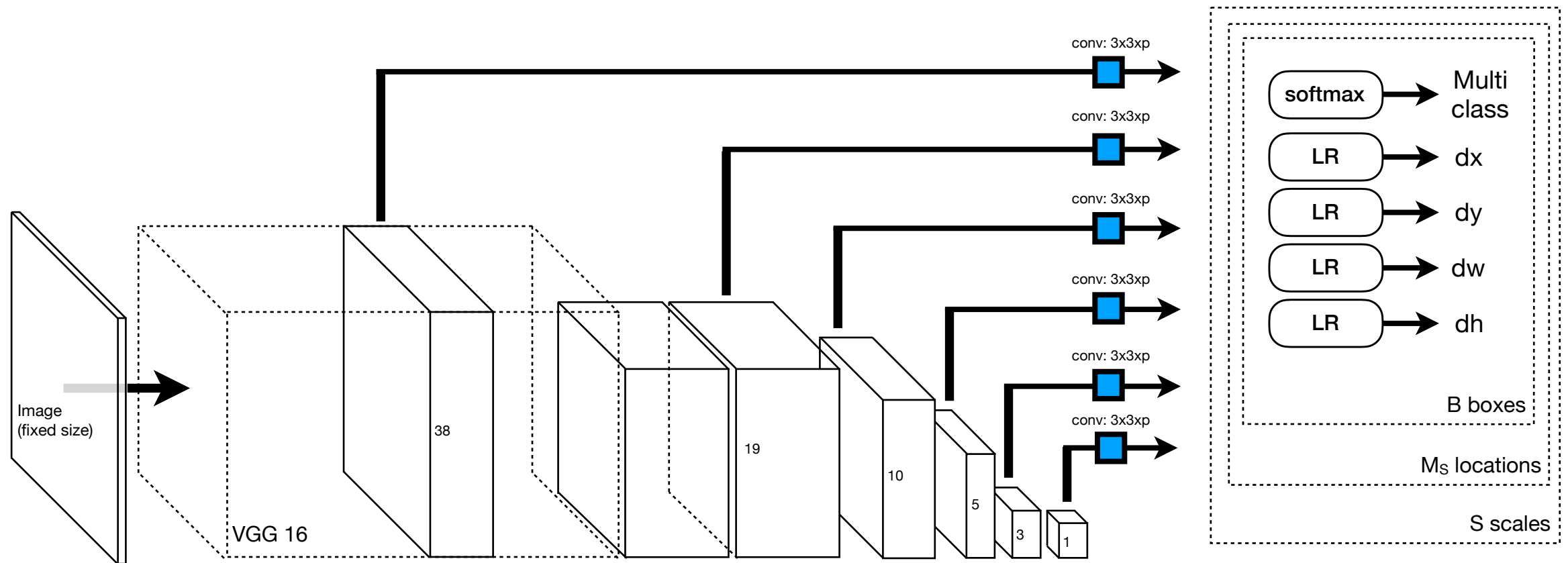
SSD

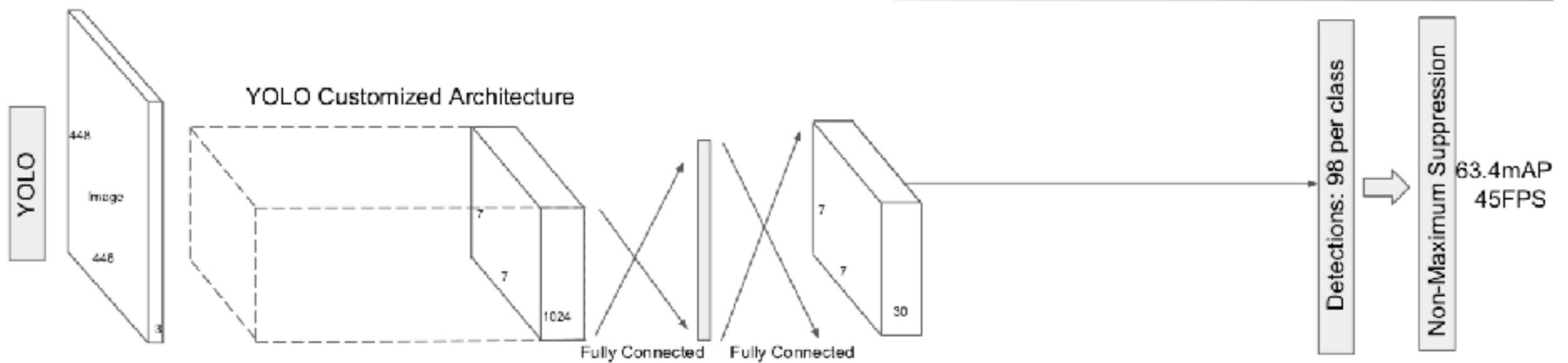
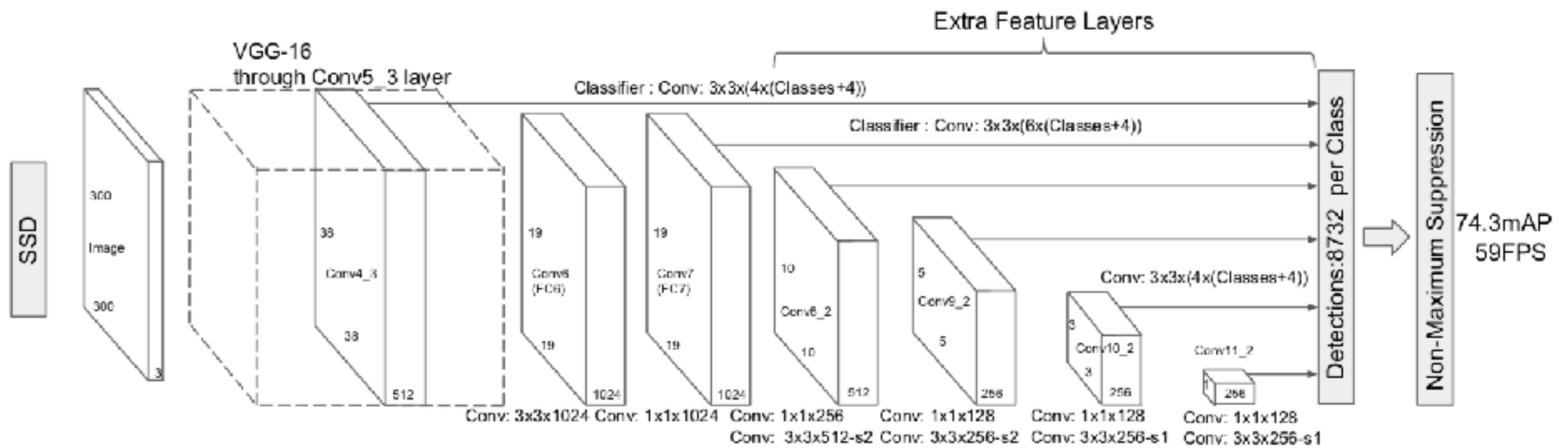
(Single Shot Multi-box Detection)

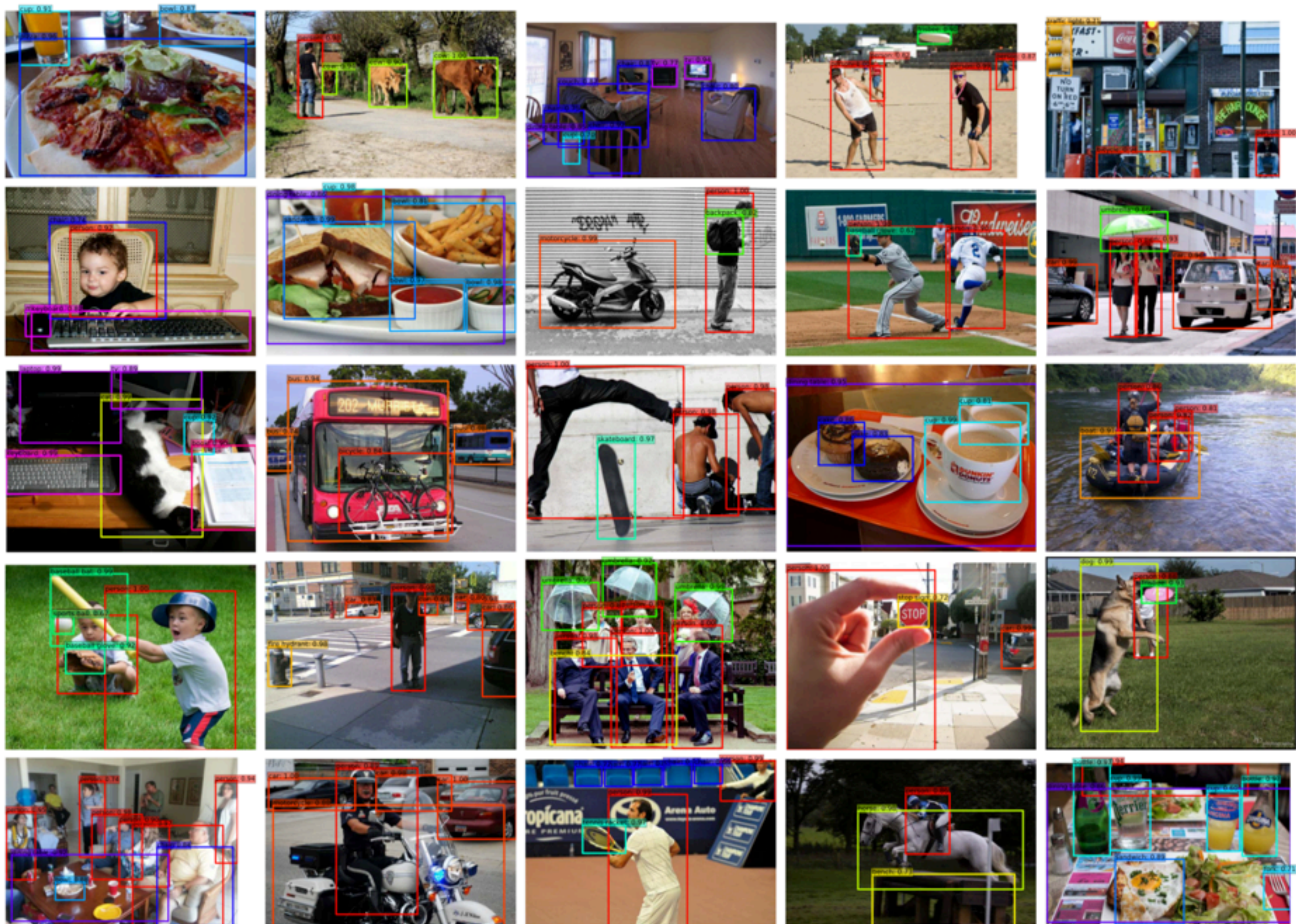


SSD

(Single Shot Multi-box Detection)

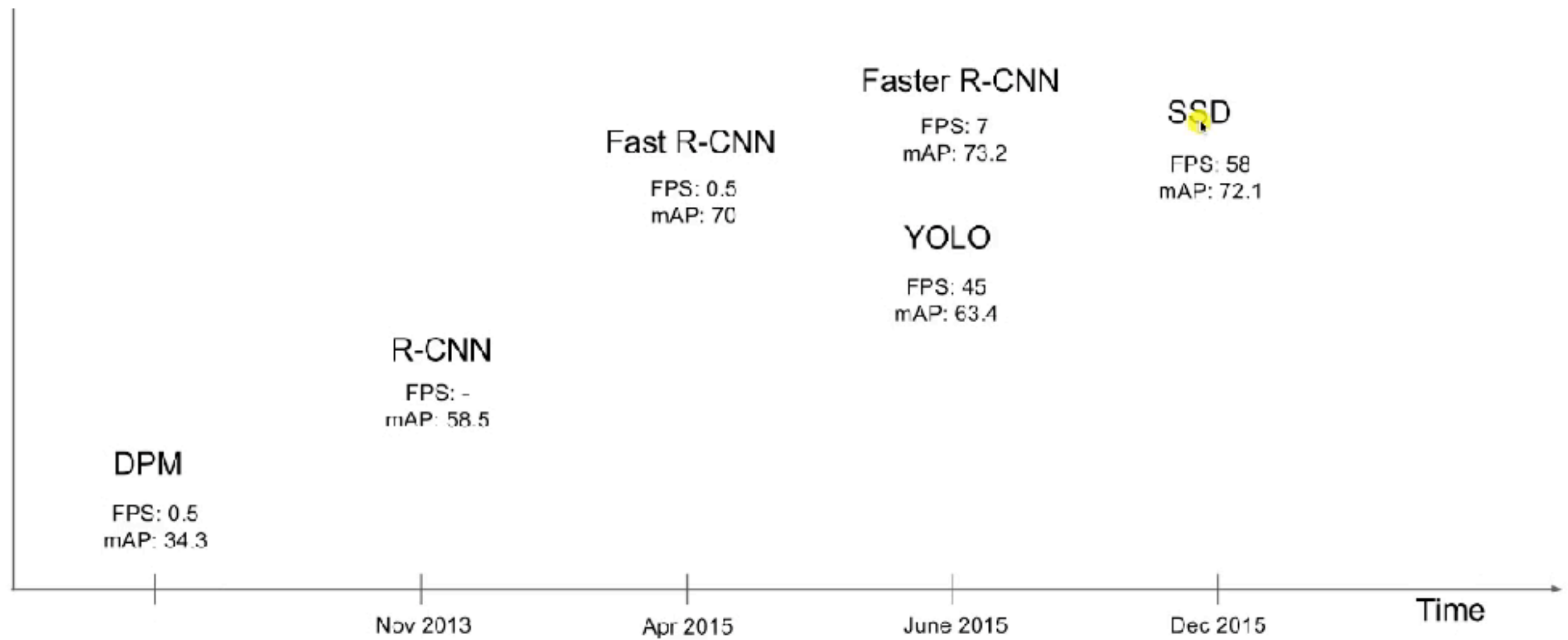






Method	mAP	FPS	batch size	# Boxes	Input resolution
Faster R-CNN (VGG16)	73.2	7	1	~ 6000	$\sim 1000 \times 600$
Fast YOLO	52.7	155	1	98	448×448
YOLO (VGG16)	66.4	21	1	98	448×448
SSD300	74.3	46	1	8732	300×300
SSD512	76.8	19	1	24564	512×512
SSD300	74.3	59	8	8732	300×300
SSD512	76.8	22	8	24564	512×512

Table 7: **Results on Pascal VOC2007 test.** SSD300 is the only real-time detection method that can achieve above 70% mAP. By using a larger input image, SSD512 outperforms all methods on accuracy while maintaining a close to real-time speed.



Important Concepts

- **Direct multi-scale regression:**
Obtain bound box information from different layers
- **Convolutional regression:**
Avoid large intermediate representations, regress values using small convolutional filters