Method	Network	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Faster [25]	VGG	73.2	76.5	79	70.9	65.5	52.1	83.1	84.7	86.4	52	81.9	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83	72.6
ION [2]	VGG	75.6	79.2	83.1	77.6	65.6	54.9	85.4	85.1	87	54.4	80.6	73.8	85.3	82.2	82.2	74.4	47.1	75.8	72.7	84.2	80.4
Faster [15]	Residual-101	76.4	79.8	80.7	76.2	68.3	55.9	85.1	85.3	89.8	56.7	87.8	69.4	88.3	88.9	80.9	78.4	41.7	78.6	79.8	85.3	72
MR-CNN [10]	VGG	78.2	80.3	84.1	78.5	70.8	68.5	88	85.9	87.8	60.3	85.2	73.7	87.2	86.5	85	76.4	48.5	76.3	75.5	85	81
R-FCN [4]	Residual-101	80.5	79.9	87.2	81.5	72	69.8	86.8	88.5	89.8	67	88.1	74.5	89.8	90.6	79.9	81.2	53.7	81.8	81.5	85.9	79.9
SSD300 [21]	VGG	77.5	79.5	83.9	76	69.6	50.5	87	85.7	88.1	60.3	81.5	77	86.1	87.5	83.9	79.4	52.3	77.9	79.5	87.6	76.8
SSD512 [21]	VGG	79.5	84.8	85.1	81.5	73	57.8	87.8	88.3	87.4	63.5	85.4	73.2	86.2	86.7	83.9	82.5	55.6	81.7	79	86.6	80
DSSD321 [8]	Residual-101	78.6	81.9	84.9	80.5	68.4	53.9	85.6	86.2	88.9	61.1	83.5	78.7	86.7	88.7	86.7	79.7	51.7	78	80.9	87.2	79.4
DSSD513 [8]	Residual-101	81.5	86.6	86.2	82.6	74.9	62.5	89	88.7	88.8	65.2	87	78.7	88.2	89	87.5	83.7	51.1	86.3	81.6	85.7	83.7
STDN300 [38]	DenseNet-169	78.1	81.1	86.9	76.4	69.2	52.4	87.7	84.2	88.3	60.2	81.3	77.6	86.6	88.9	87.8	76.8	51.8	78.4	81.3	87.5	77.8
STDN321 [38]	DenseNet-169	79.3	81.2	88.3	78.1	72.2	54.3	87.6	86.5	88.8	63.5	83.2	79.4	86.1	89.3	88.0	77.3	52.5	80.3	80.8	86.3	82.1
STDN513 [38]	DenseNet-169	80.9	86.1	89.3	79.5	74.3	61.9	88.5	88.3	89.4	67.4	86.5	79.5	86.4	89.2	88.5	79.3	53.0	77.9	81.4	86.6	85.5
VDNet	Resnet-101	**86.2**	95.8	98.1	98.4	65.1	94.6	90.1	96.2	71.7	72.3	54.6	97.9	95.6	89.2	90.1	93.2	69.1	89.2	82.1	93.4.6	74.0

Table 1. PASCAL VOC 2007 Test Detection Results. Note that the minimum dimension of the input image for Faster and R-FCN is 600, and the speed is less than 10 frames per second. SSD300 indicates the input image dimension of SSD is 300×300. Large input sizes can lead to better results, but this increases running times. All models were trained on the union of the trainval set from VOC 2007 and VOC 2012 and tested on the VOC 2007 test set.