

1. Caltech Experiment Results

The details of the experiment results on Caltech pedestrian detection are shown in this section. As described in the experiment part of our paper, we use R-FCN detection framework and modified Inception network architectures. Figure 1 shows the R-FCN detection results of different models trained by mimic method, fine-tuning from ImageNet pre-trained model and training from scratch.

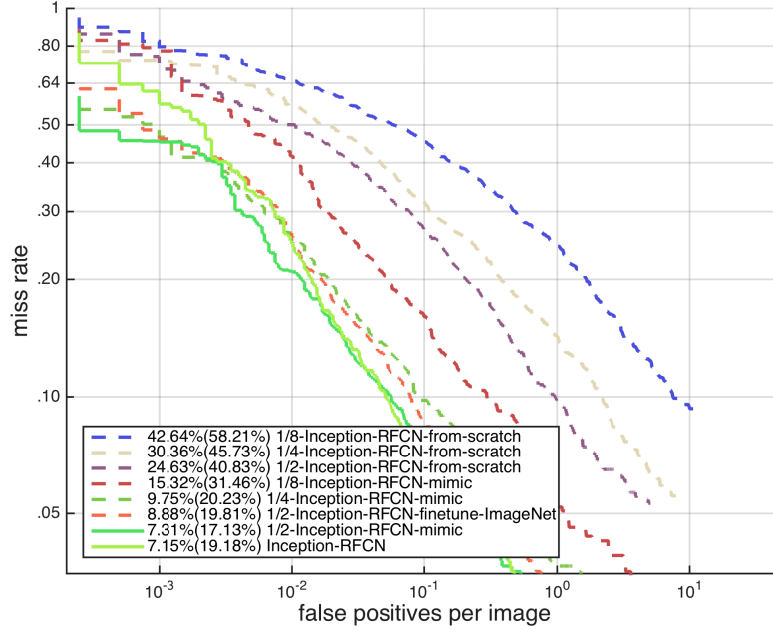


Figure 1: Comparison of R-FCN detection results evaluated on Caltech with the new annotations ($MR_2(MR_4)$).

2. PASCAL VOC Experiment Results

We present details of our experiment results evaluated on PASCAL VOC 2007. We train Faster R-CNN jointly on the union set of PASCAL VOC 2007 trainval and 2012 trainval. We use at most 300 proposals for each image during testing. The detailed results are shown in Table 1. Models trained by mimic method outperform those fine-tuned from ImageNet pre-trained model or trained from scratch.

method	mAP	areo	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
<i>Inception</i> ¹	75.7	80.0	84.2	76.0	69.4	57.6	84.9	85.1	88.0	50.6	79.0	67.7	85.7	85.7	82.0	80.3	45.0	76.9	73.4	86.6	76.1
<i>1/2Inception</i> ¹	72.4	75.4	84.3	72.9	62.2	49.7	83.2	82.7	86.8	44.1	75.7	64.8	83.0	86.0	81.4	78.6	37.8	73.3	70.1	82.0	73.5
<i>1/2Inception</i> ²	72.8	77.2	81.0	69.5	65.9	51.8	83.7	84.0	84.9	47.3	75.5	66.0	82.0	85.2	79.7	79.2	42.3	72.8	69.8	84.1	73.3
<i>1/4Inception</i> ³	42.1	51.1	49.0	27.6	30.2	8.4	54.1	60.6	43.8	18.3	41.0	44.6	37.0	63.9	57.3	53.0	11.7	40.0	42.4	58.3	49.6
<i>1/4Inception</i> ²	65.8	72.0	77.0	59.4	58.1	38.3	78.5	80.0	76.3	40.0	65.3	57.8	71.0	79.8	75.9	75.4	37.1	64.9	63.1	79.6	65.8
<i>1/4Inception</i> ⁴	67.7	71.0	77.1	63.9	60.4	40.9	79.3	81.1	79.8	41.3	71.5	57.1	76.0	82.9	76.4	74.5	37.2	66.2	66.7	82.7	67.1
<i>1/8Inception</i> ³	34.8	46.3	42.4	21.2	26.1	4.4	41.6	53.1	37.6	13.4	32.7	34.0	30.7	56.6	50.7	44.1	5.7	33.0	33.1	49.0	39.6
<i>1/8Inception</i> ²	53.8	61.2	62.2	43.3	49.1	17.5	65.4	72.9	61.3	27.7	48.4	56.6	53.7	72.2	65.0	68.2	22.4	48.2	57.3	68.5	55.0
<i>1/4Inception</i> ⁴	56.1	62.7	64.3	44.7	45.2	20.5	68.7	72.2	70.1	31.7	57.0	52.9	61.6	74.0	67.4	65.8	25.9	52.0	57.0	72.3	57.1

Table 1: Results on PASCAL VOC 2007 test set with Faster R-CNN detectors by different training methods, *model*¹ stands for fine-tuning from ImageNet pre-trained model, *model*² stands for mimic+finetune, *model*³ stands for training from scratch, and *model*⁴ stands for two stage mimic method.