Supplementary Material for Paper 2003: Region Ranking SVMs for Image Classification

Anonymous CVPR submission

Paper ID 2003

1. Introduction

In this document we first detail the structure of the three-layer network used to calibrate the output of 1000 binary Region Ranking SVMs(RRSVM) that were individually trained for each object category. Second, we show more experiments that have been done on PASACL VOC 2012 Action Classification to demonstrate that: (i) RRSVM outperforms the baseline methods independent of the ConvNets; (ii) RRSVM also benefits from multiple ConvNet fusions. Third, we analyze the region-score vector distribution of each category in PASCAL VOC 2007 Image Classification dataset and show that the number of effective regions for each category is no more than 30, which reveals the fact that our proposed RRSVM effectively finds the semantical meaningful regions of objects. We finish our document by showing more qualitative results on PASCAL VOC 2007 Image Classification, PASCAL VOC 2012 Action Classification and ILSVRC 2014 Image Classification datasets.

2. Multiple RRSVMs calibration

As briefly described in Section 4.4.1, since RRSVMs are independently trained for each class, they must be calibrated to be used for multi-class classification. We use a three-layer neural network to reconcile the outputs of the 1000 classifiers. As illustrated in Fig.1, the first layer of the network is a fully connected layer with 1000 units followed by a rectified linear unit (ReLU). The second layer is the same as the first layer but without ReLU. The last layer is a soft-max layer with 1000 outputs for 1000 classes. The network is trained to minimize the negative log likelihood using all training examples.

The parameters for training this network are as follows: the two fully connected layers are initialized as identity matrices. The weight decay vector is set to 0 and the other parameters are set the same as the ones in training VGG16 Net.

Normally the network takes approximately 20 epochs to converge.

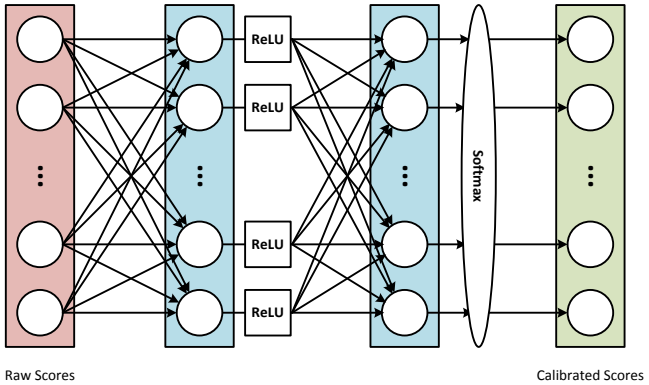

Figure 1: Three-Layer NN structure. The layers in purple are fully connected layers.

Model	jump.	phone.	playinstr.	read.	ridebike	ridehorse	run	takephoto	usecomp.	walk.	mAP
VGG16-LSSVM	85.5	64.5	95.2	71.6	93.0	95.9	82.5	69.0	90.7	62.0	81.0
VGG16-RRSVM	87.2	72.3	96.2	76.9	94.6	96.8	87.8	74.4	92.9	62.6	84.2
VGG19-LSSVM	85.5	64.5	95.4	72.4	93.0	96.0	83.4	70.9	90.7	62.1	81.4
VGG19-RRSVM	86.5	74.4	96.1	76.8	94.8	96.8	87.6	76.7	92.5	61.7	84.4
VGG19&VGG16 - RRSVM	**87.5**	**75.0**	**96.4**	**77.8**	**95.0**	**96.9**	**88.4**	**77.9**	**92.9**	**63.7**	**85.1**

Table 1: **Average precision (%) on VOC2012 Action validation set.** Note that: (i) RRSVM outperforms LSSVM by a large margin using any VGG net; (ii) RRSVM using VGG16 and VGG19 outperforms both RRSVMs using single net.

3. Additional Experiments on PASCAL VOC 2012 Action Classification

For PASCAL VOC 2012 Action Classification dataset, in addition to the experiments listed in the paper, we also tested the performance of RRSVM using VGG16 and VGG19 as feature extractor separately. The results of these experiments on validation set in Tab. 1 showed that our proposed RRSVM outperforms LSSVM on both settings that uses individual VGG16 or VGG19 for feature extraction. Also, VGG16 and VGG19 have comparable performances, but simply averaging them further improves the results.

4. Region-Score Vector analysis for each category in PASCAL VOC 2007

We visualized the Region-Score Vector values for each category in PASCAL VOC 2007 in Fig. 2. The distributions show that for all the categories, less than 25 regions are considered to be important for object classification tasks. The results also demonstrate the inadequacy of simply using max-pooling or sum-pooling for object classification tasks.

5. Qualitative results

5.1. PASCAL VOC 2007 Image Classification dataset

Some visualizations of the test results are shown in Figures 3–6. Note that the our proposed RRSVM successfully finds the semantically important regions(blue boxes) for classifying the target objects in yellow bounding boxes.

5.2. PASCAL VOC 2012 Action Classification dataset

Visualizations of some examples in PASCAL VOC 2012 validation set are shown in Fig.7-16. Note that for each of the 10 actions, our proposed RRSVM snaps the regions that either best represents the target action or best distinguishes the target action from others.

5.3. ILSVRC 2014 Image Classification

In this section, we first of all show categories where in the validation set: (1) RRSVM outperforms the baseline most (+13%, n03673021, “liner”, as shown in Fig. 17); (2) RRSVM underperforms the baseline most (-23.7%, n02447721, “gong”, as shown in Fig.18).

We further randomly select more examples from the validation set and show them in Figures 19-23. Based on these examples we observe that RRSVM successfully extracted the semantically meaningful regions for all the categories.

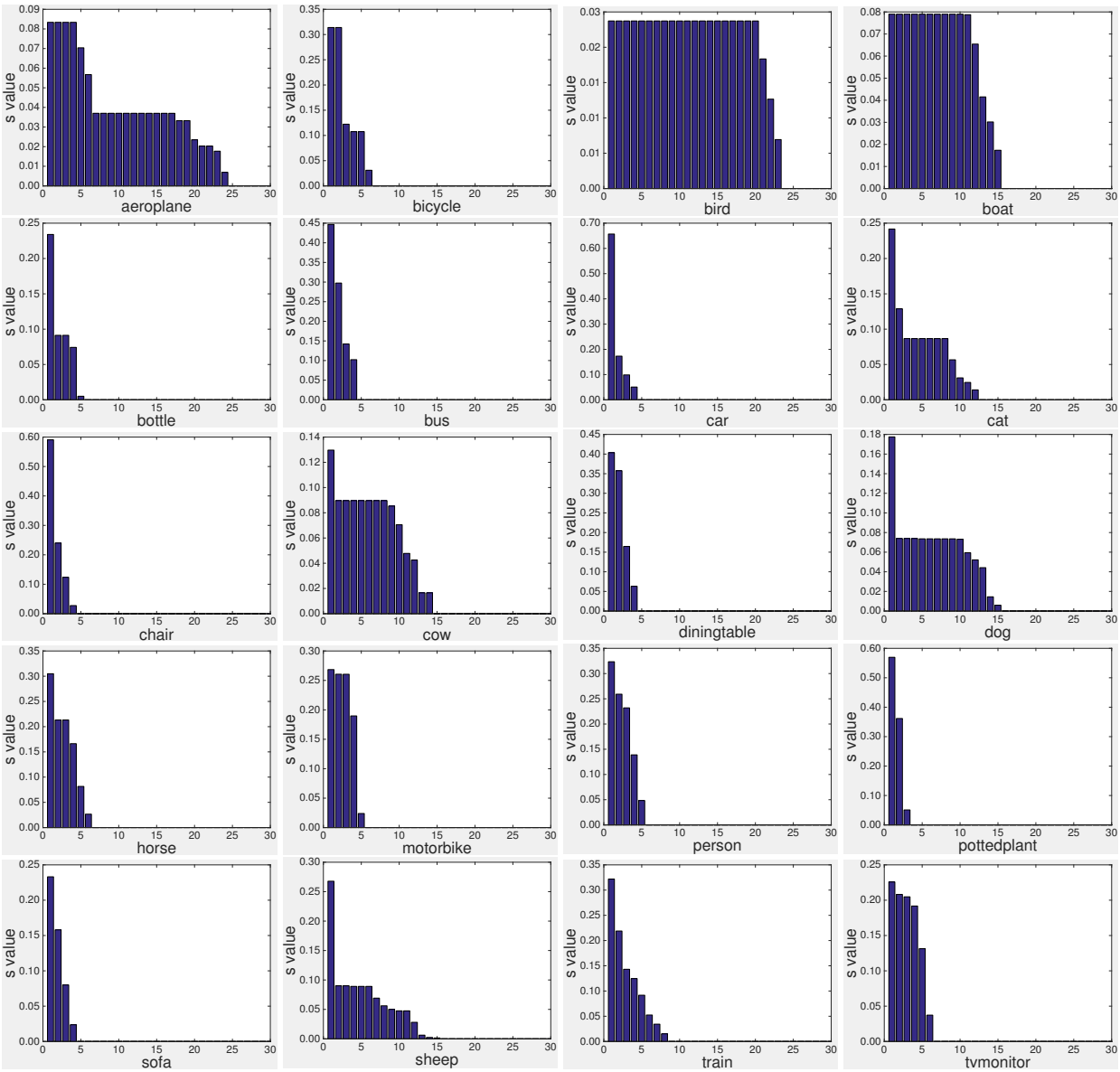

Figure 2: Distribution of region-score vectors for 20 categories in PASCAL VOC 2007 Image Classification dataset. Note that the effect number of regions for each category demonstrates the inadequacy of max-pooling or sum-pooling.

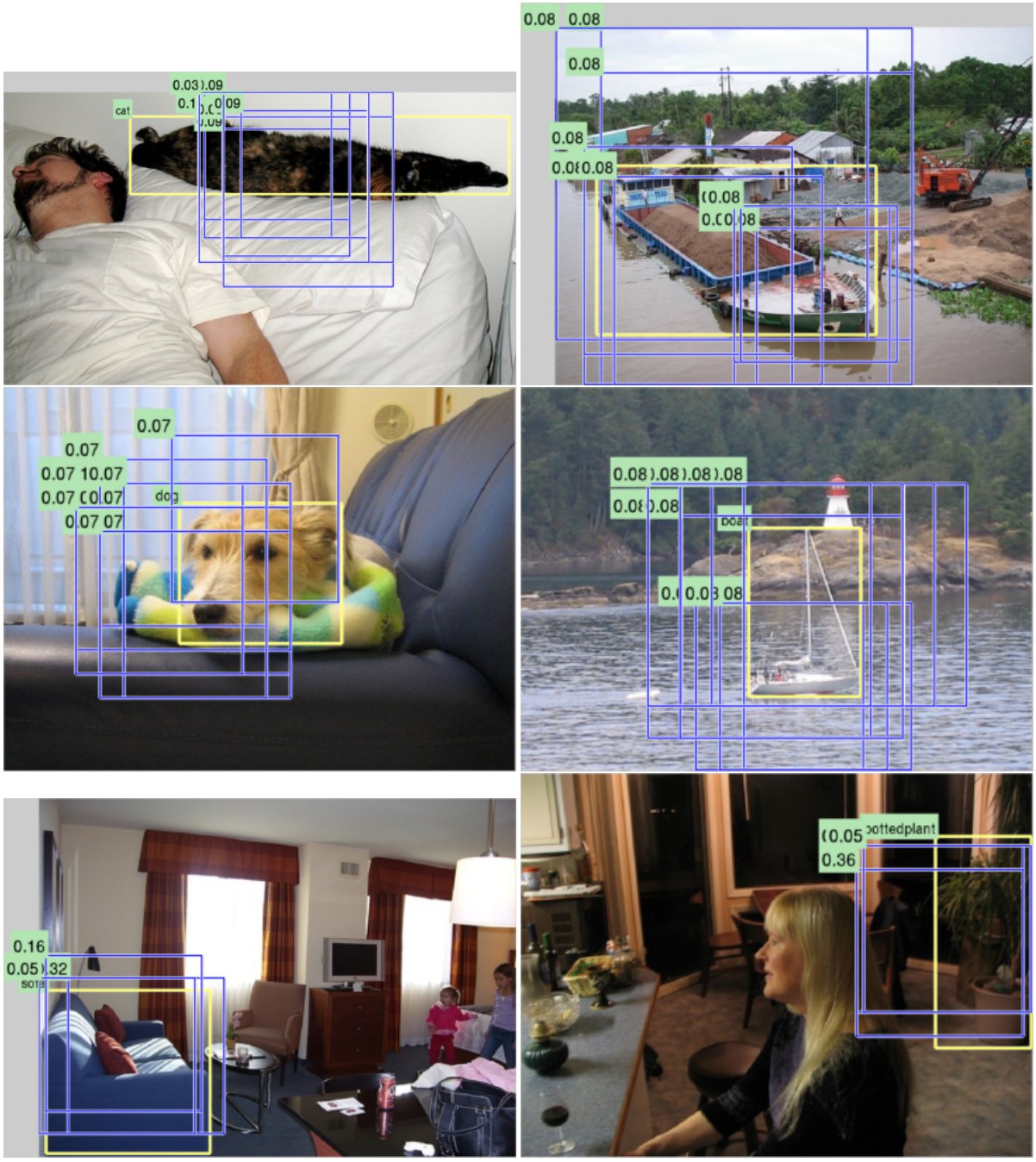

Figure 3: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific category. Best viewed on a digital device.

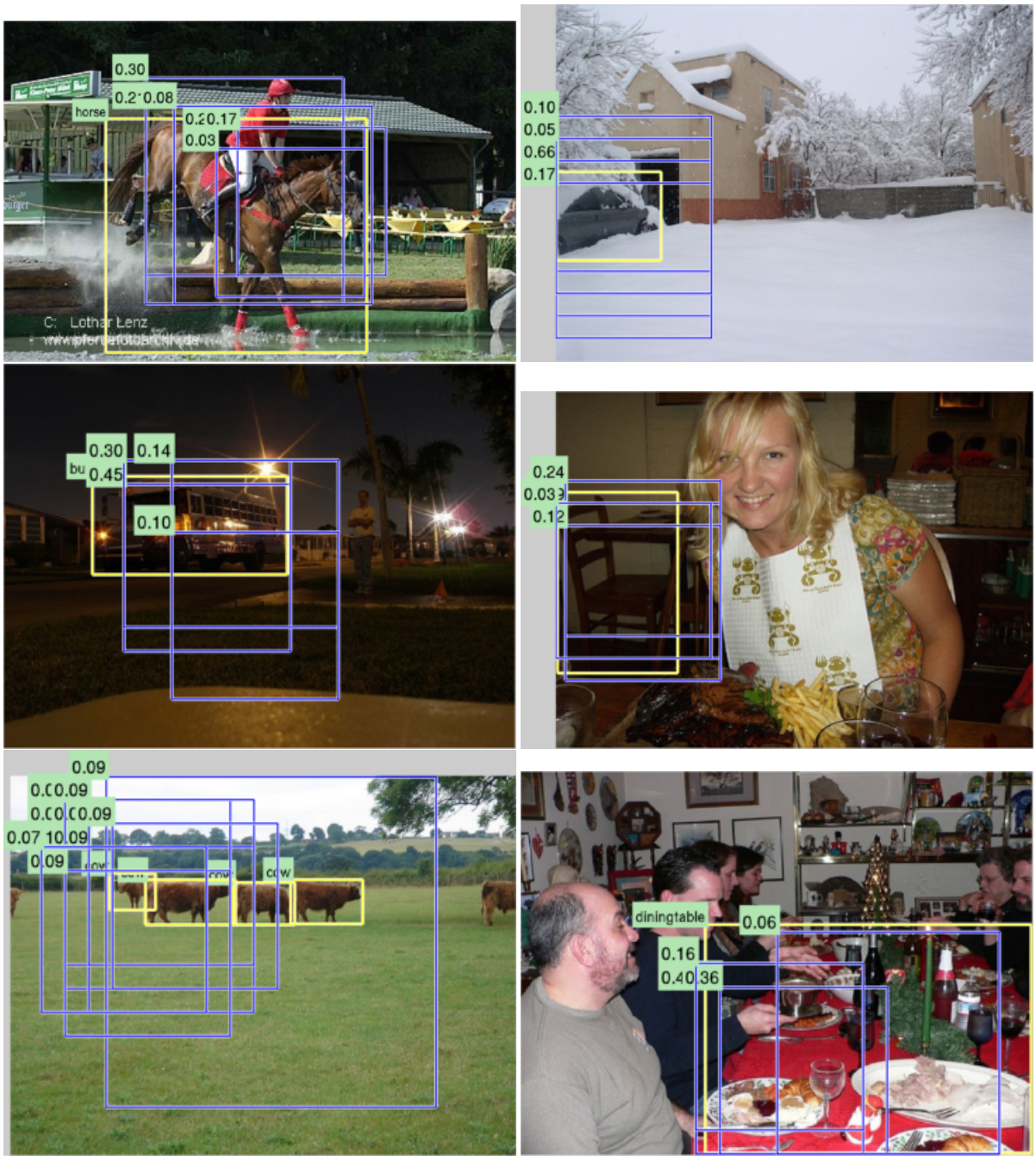

Figure 4: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific category. Best viewed on a digital device.

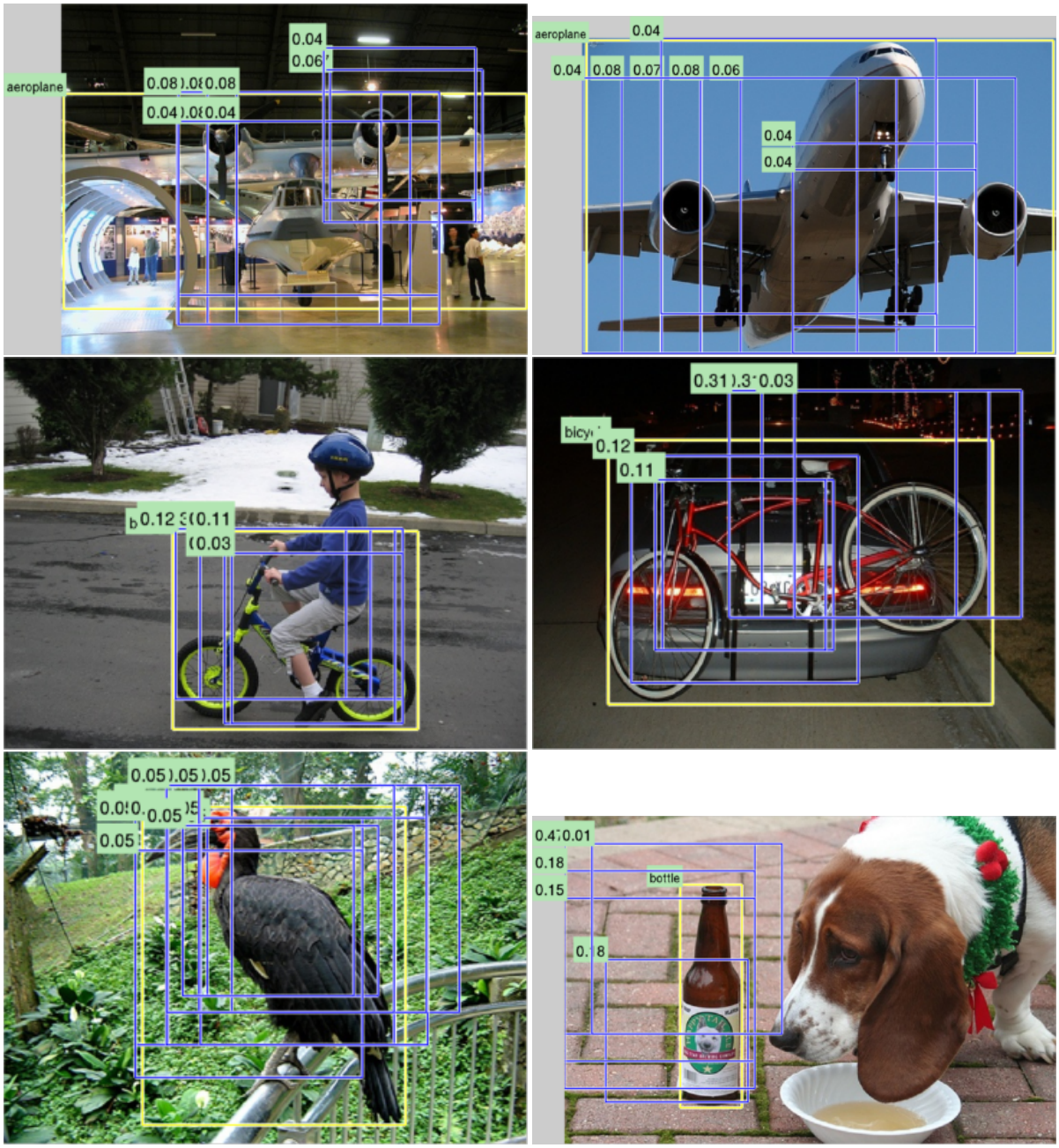

Figure 5: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific category. Best viewed on a digital device.

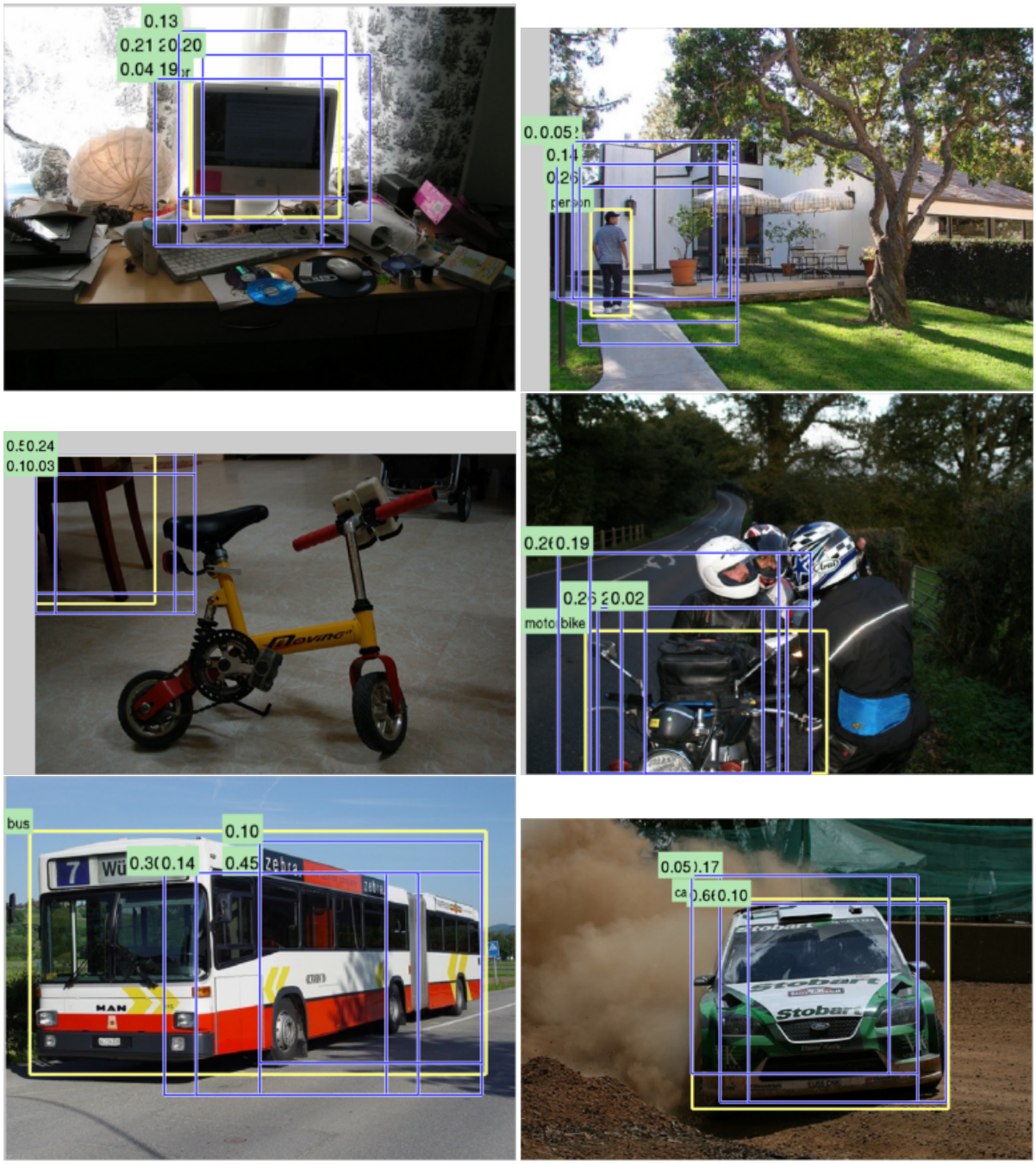

Figure 6: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific category. Best viewed on a digital device.

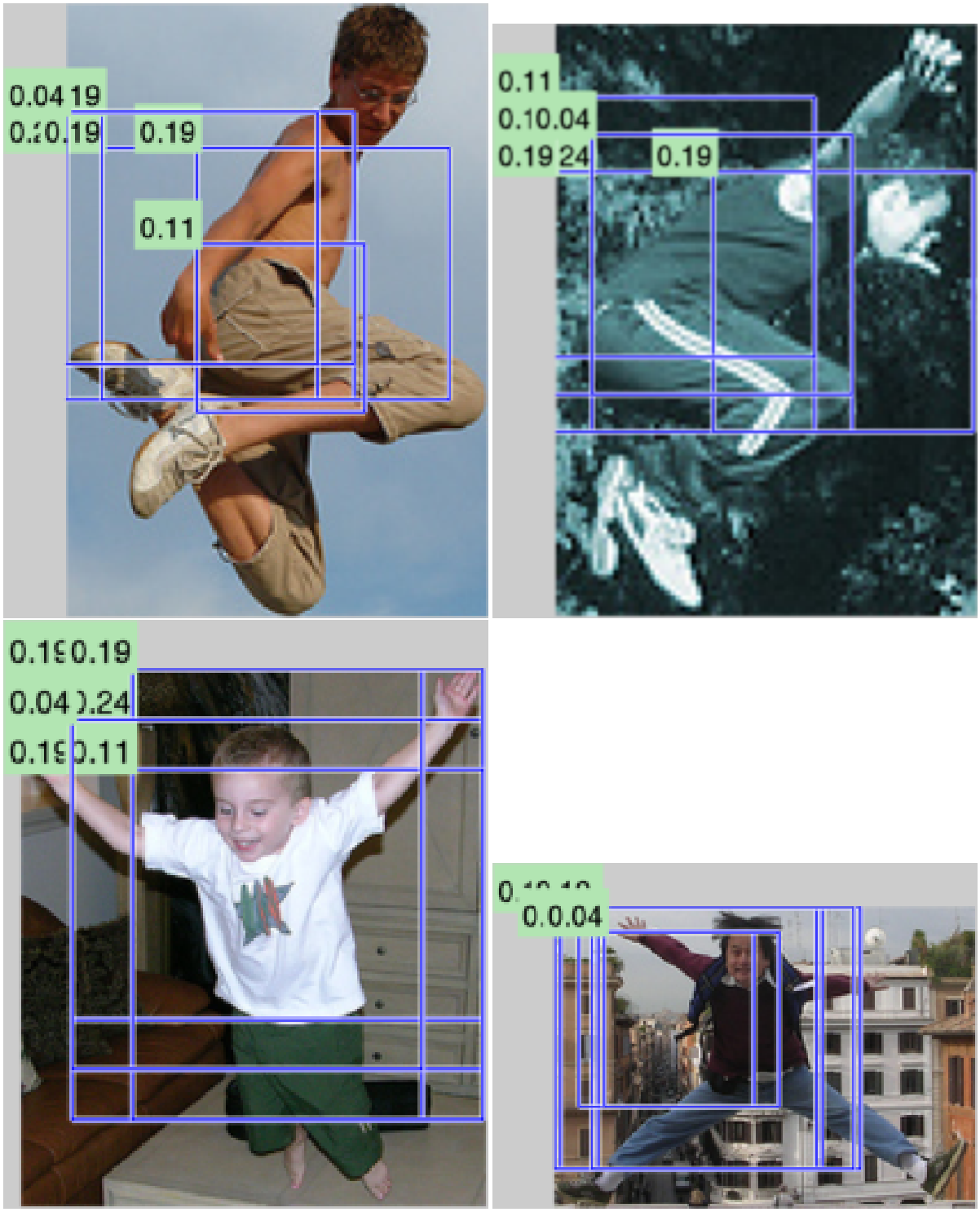

Figure 7: Examples of **jumping** in PASCAL VOC 2012 Action Classification.

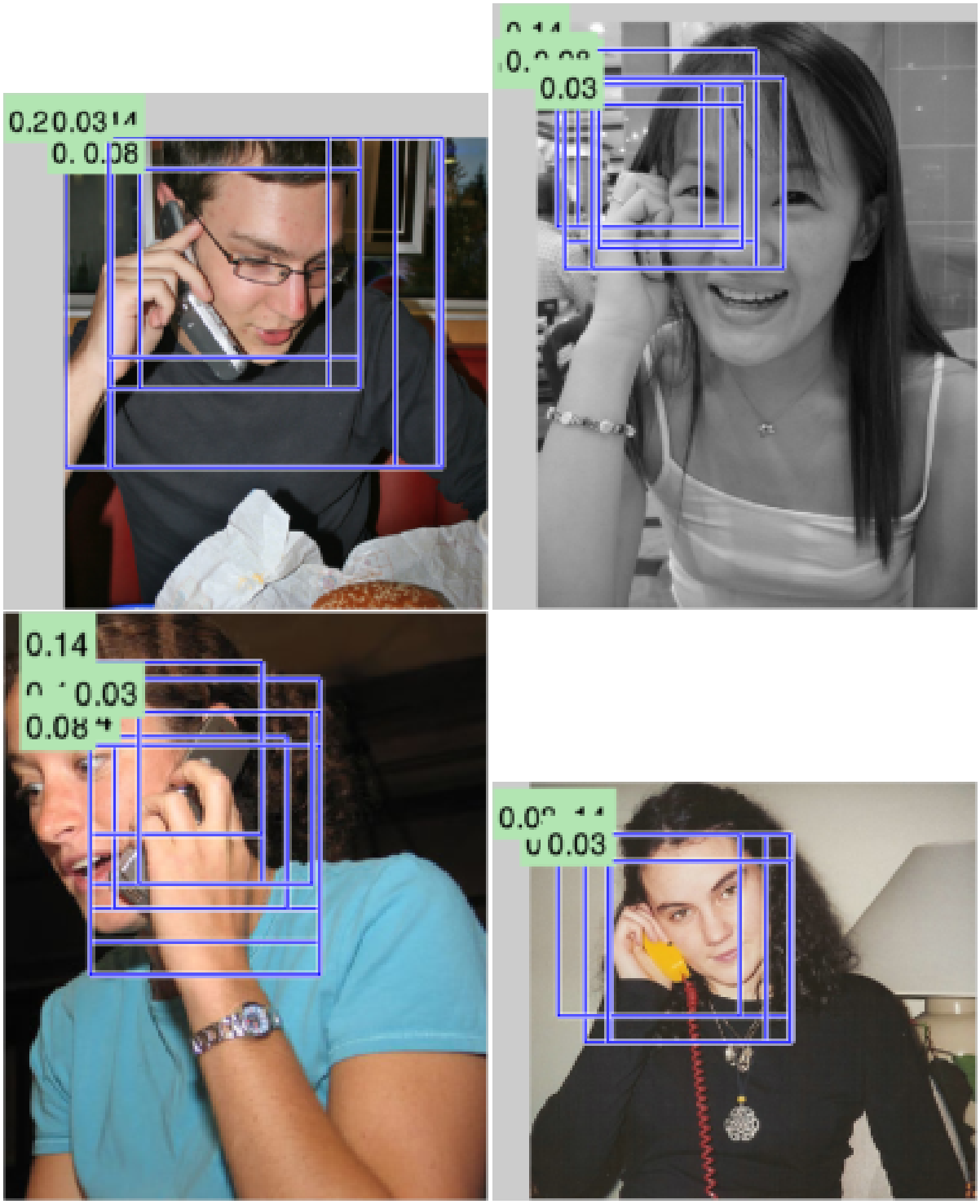

Figure 8: Examples of **phoning** in PASCAL VOC 2012 Action Classification.

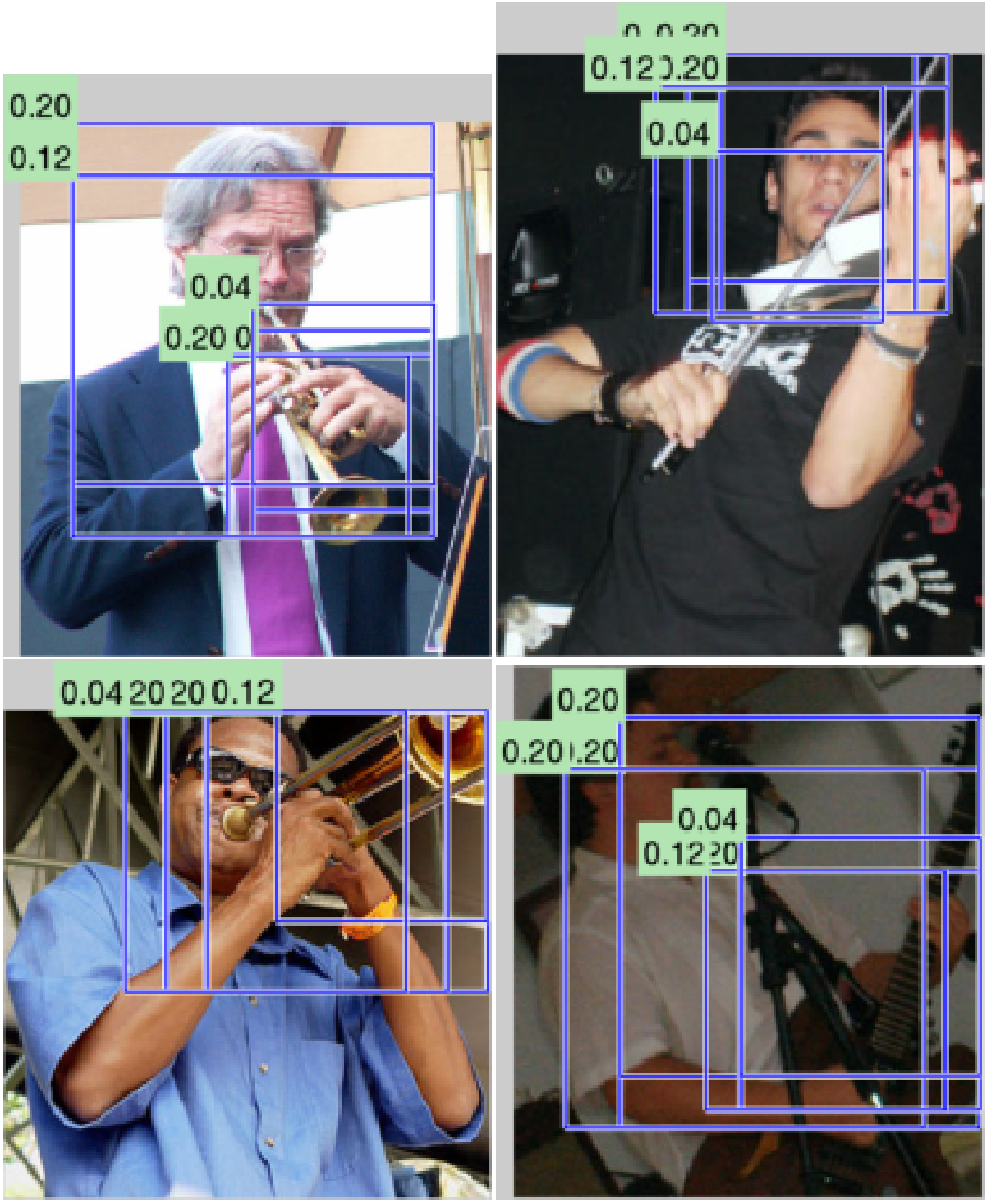

Figure 9: Examples of **playinginstrument** in PASCAL VOC 2012 Action Classification.

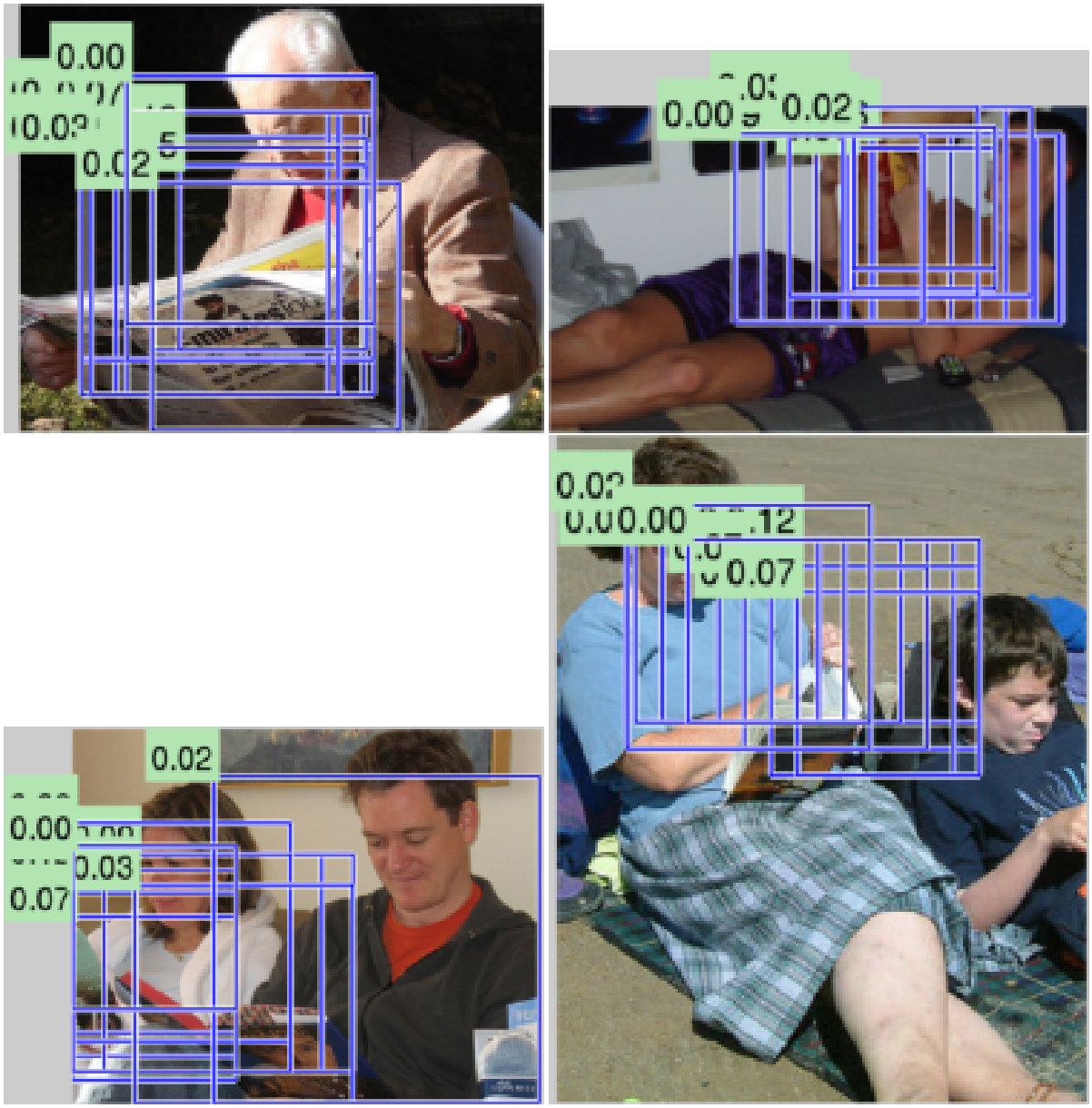

Figure 10: Examples of **reading** in PASCAL VOC 2012 Action Classification.

Figure 11: Examples of **ridingbike** in PASCAL VOC 2012 Action Classification.

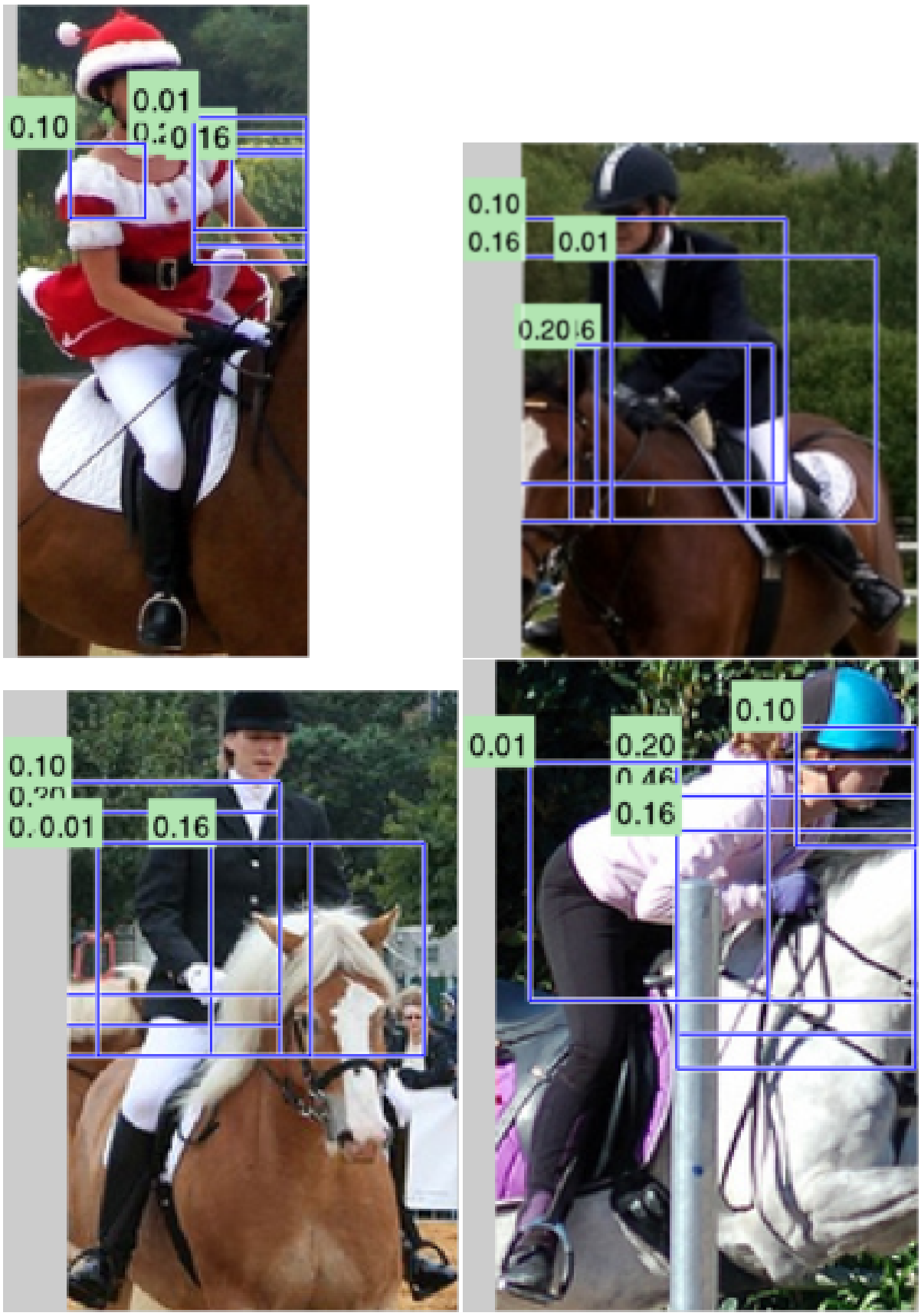

Figure 12: Examples of **ridinghorse** in PASCAL VOC 2012 Action Classification.

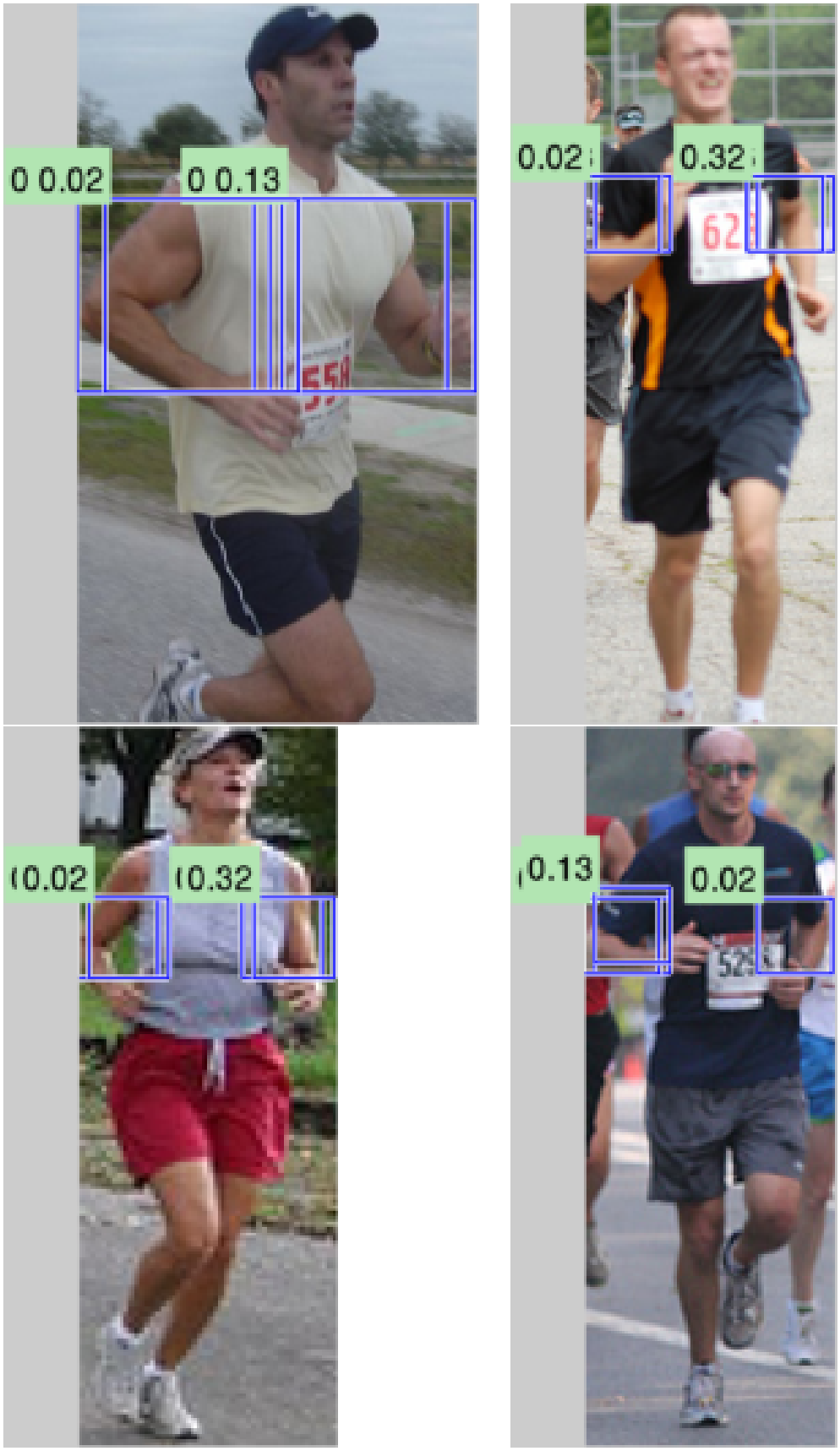

Figure 13: Examples of **running** in PASCAL VOC 2012 Action Classification.

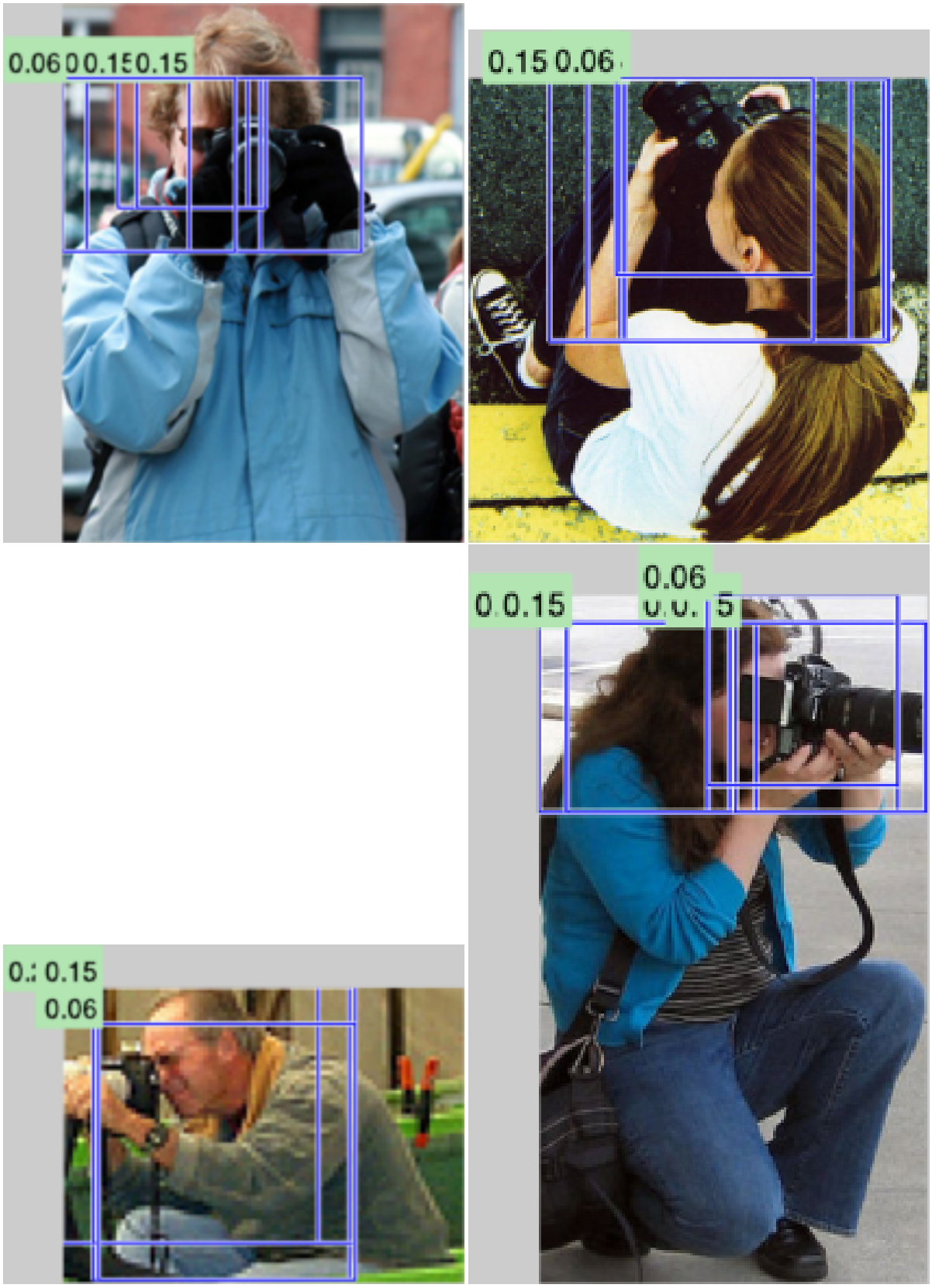

Figure 14: Examples of **takingphoto** in PASCAL VOC 2012 Action Classification.

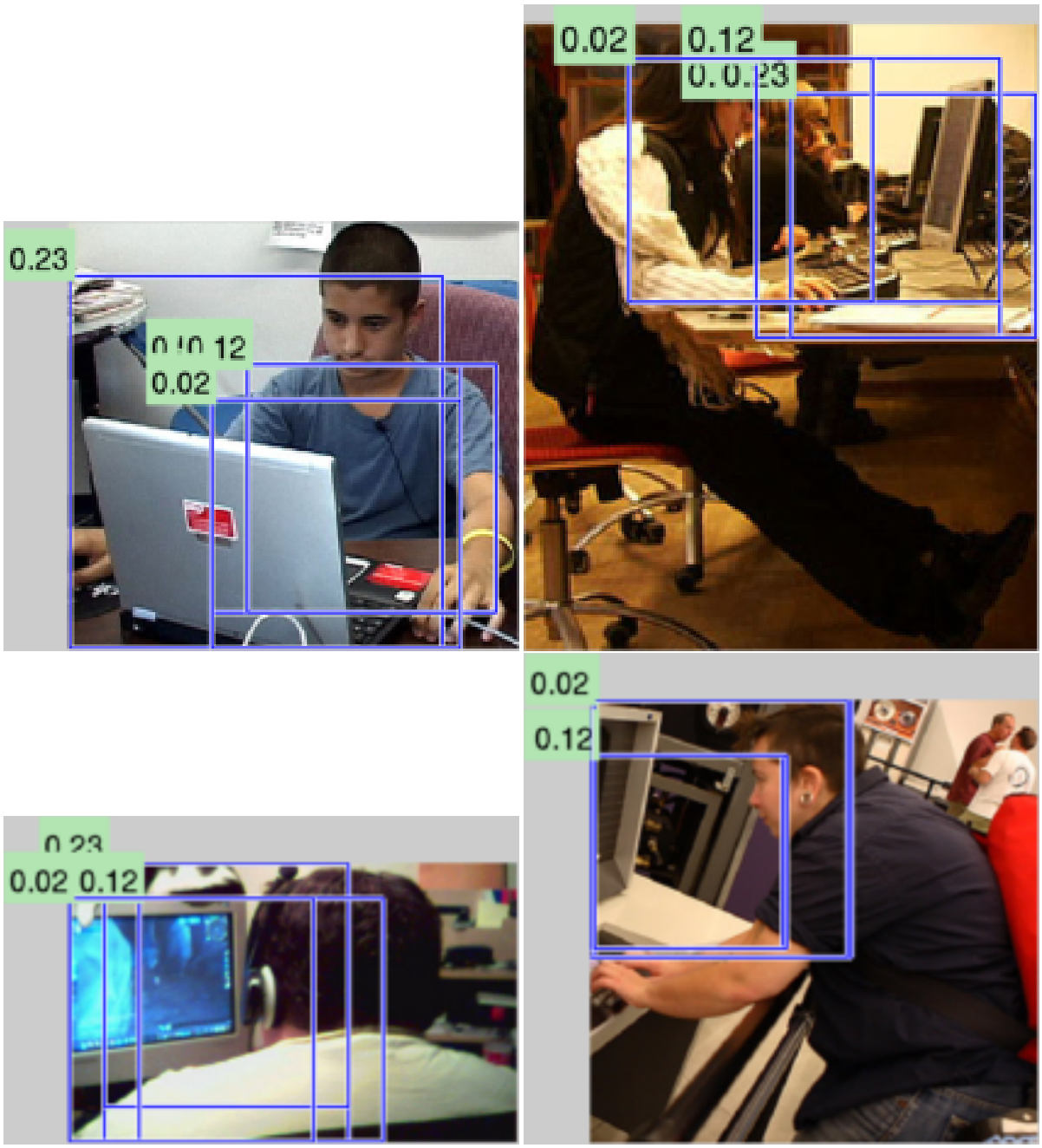

Figure 15: Examples of **usingcomputer** in PASCAL VOC 2012 Action Classification.

Figure 16: Examples of **walking** in PASCAL VOC 2012 Action Classification.

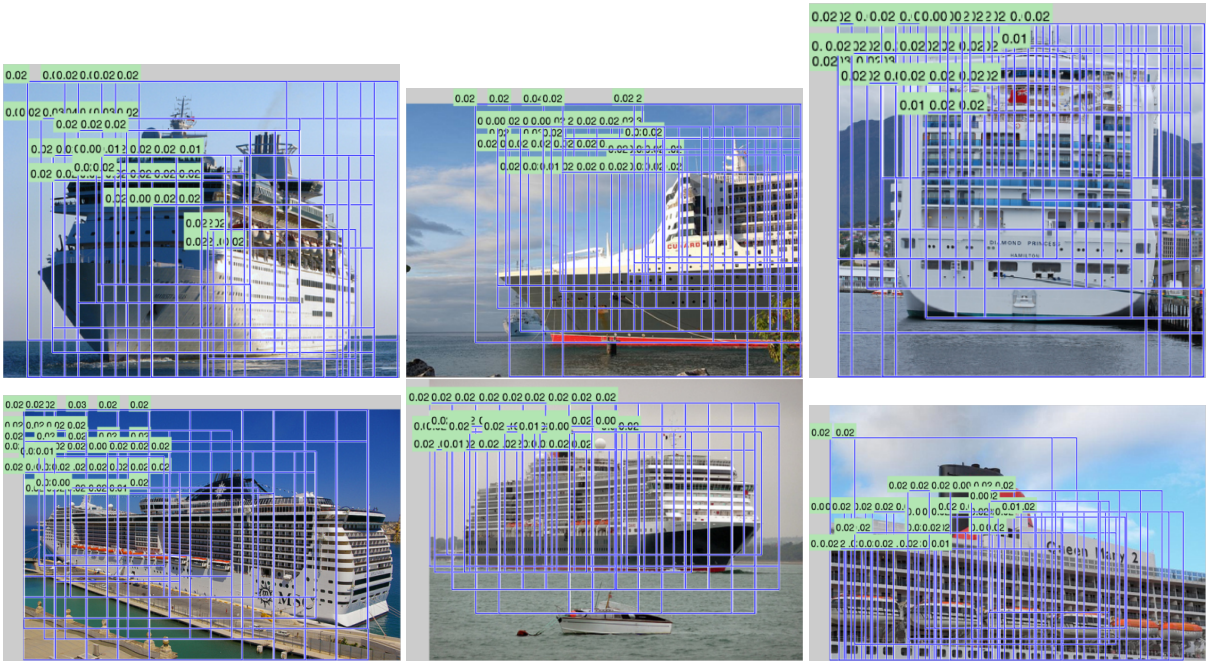

Figure 17: Examples in category n03673021(liner) that has the largest improvement over baseline. Note that the liners have relative stable texture and color, which makes the base classifier very stable to detect regions that belong to the liner.

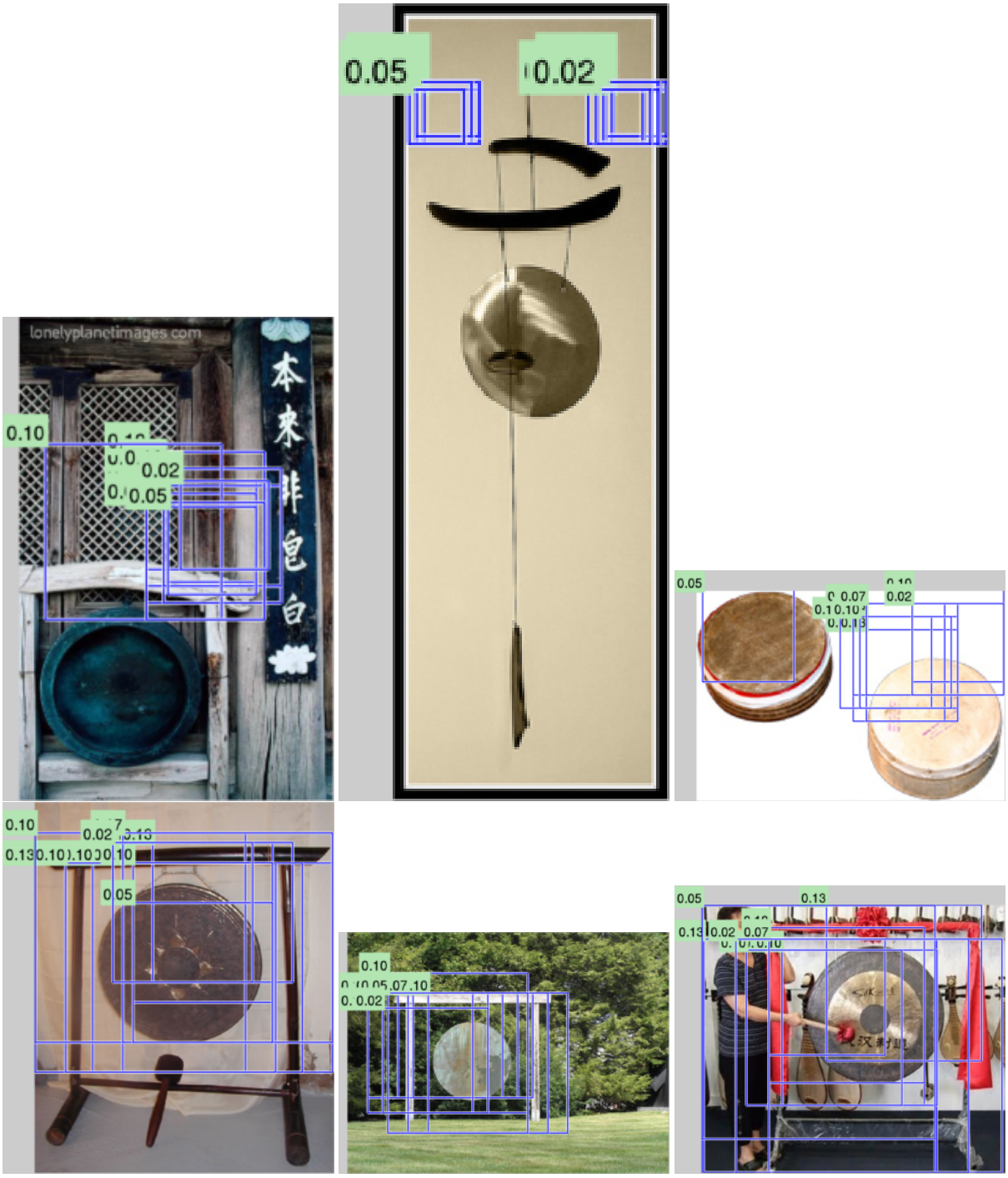

Figure 18: Examples in category n02447721(gong) that has the largest drop compared to baseline.Top row shows the cases that are wrongly classified. Notice the shape and color variance of objects.

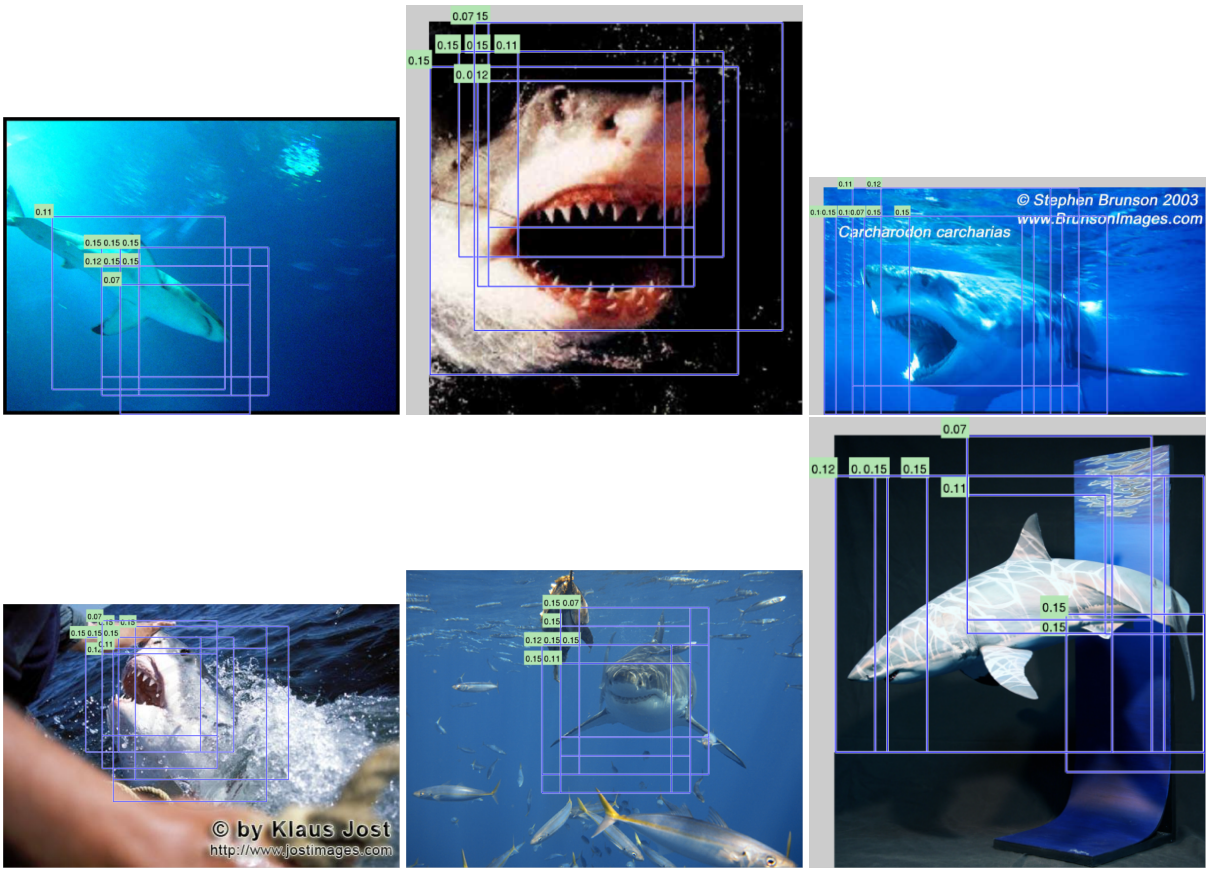

Figure 19: More examples of ILSVRC2014: n01484850 (white shark).

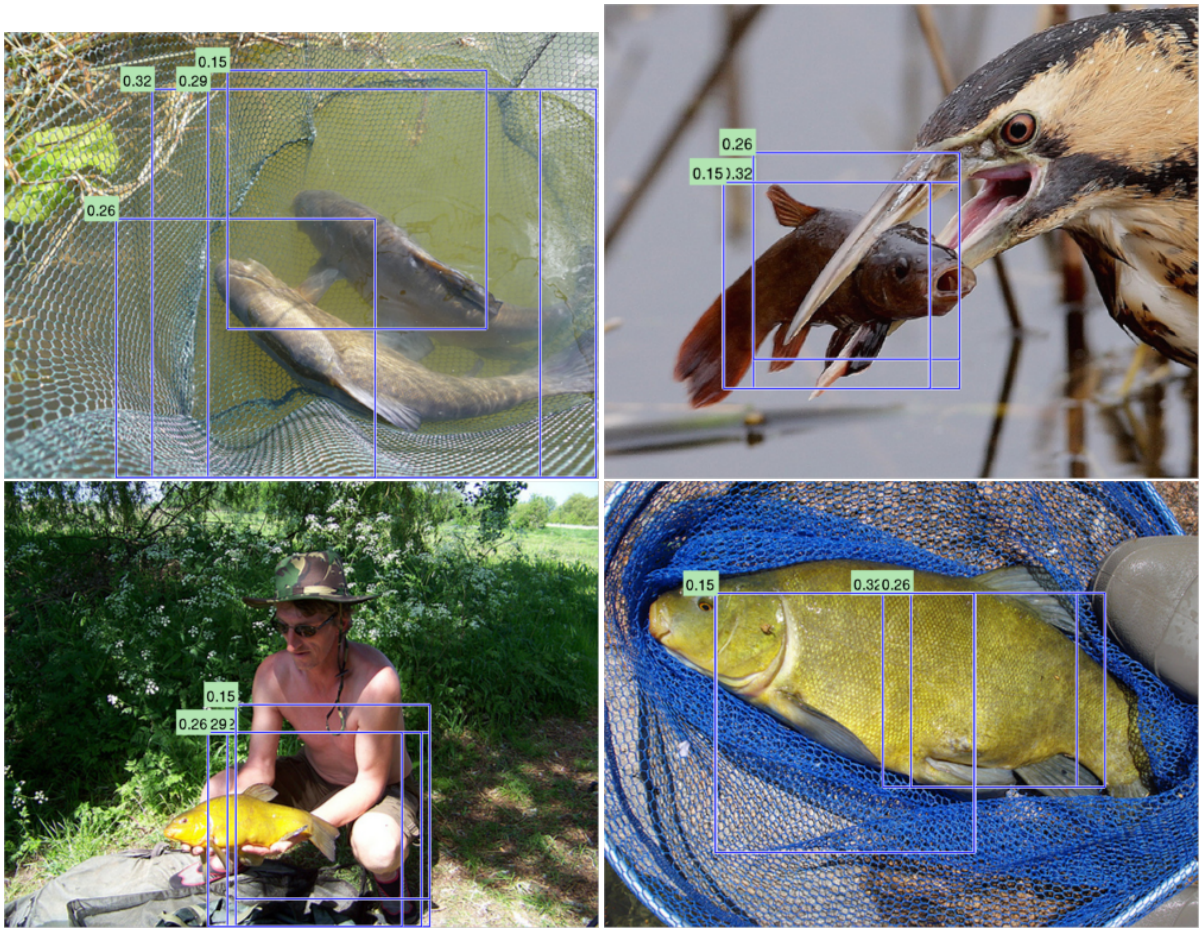

Figure 20: More examples of ILSVRC2014: n01440764(tench).

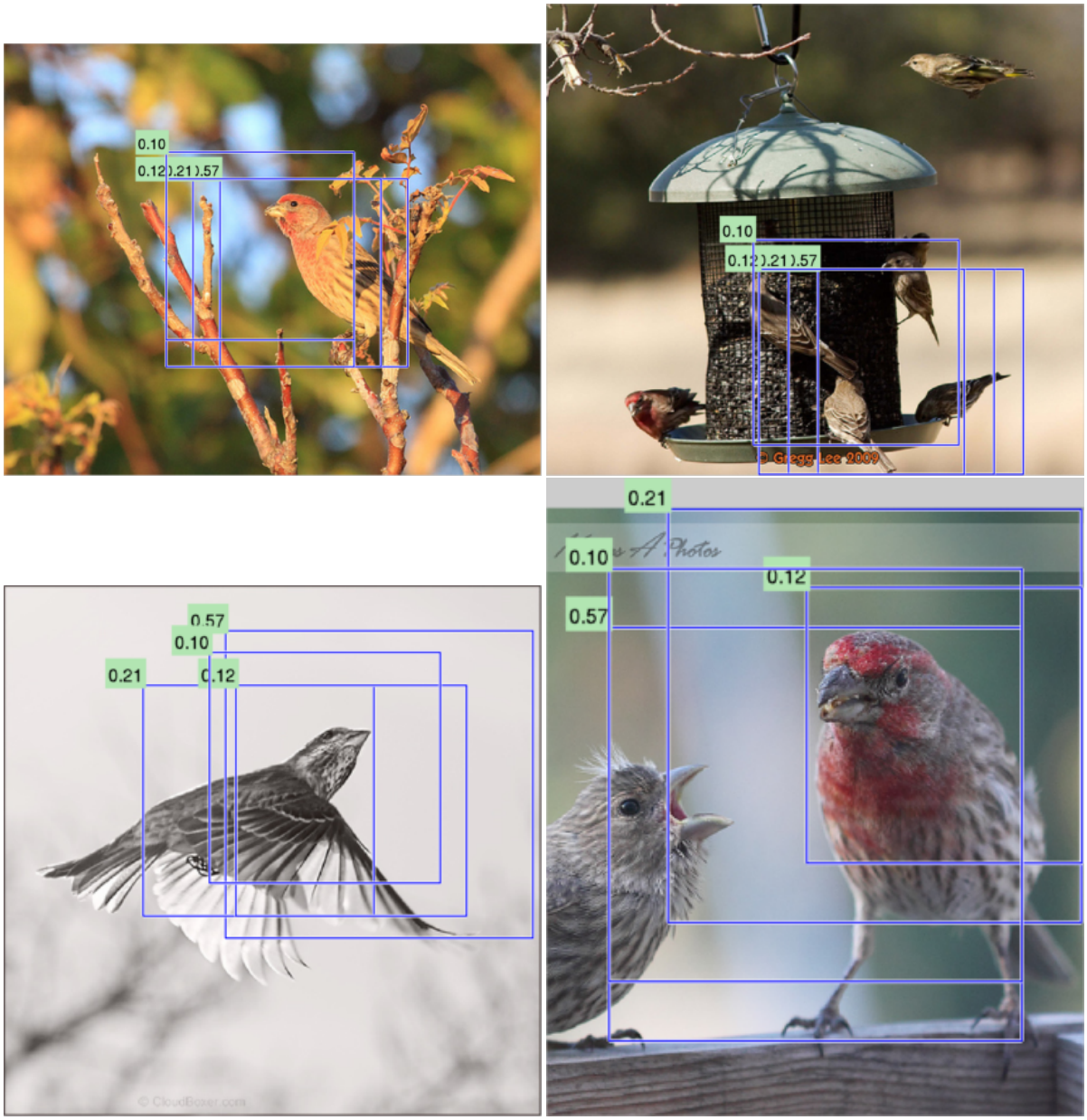

Figure 21: More examples of ILSVRC2014: n01532829(house finch).

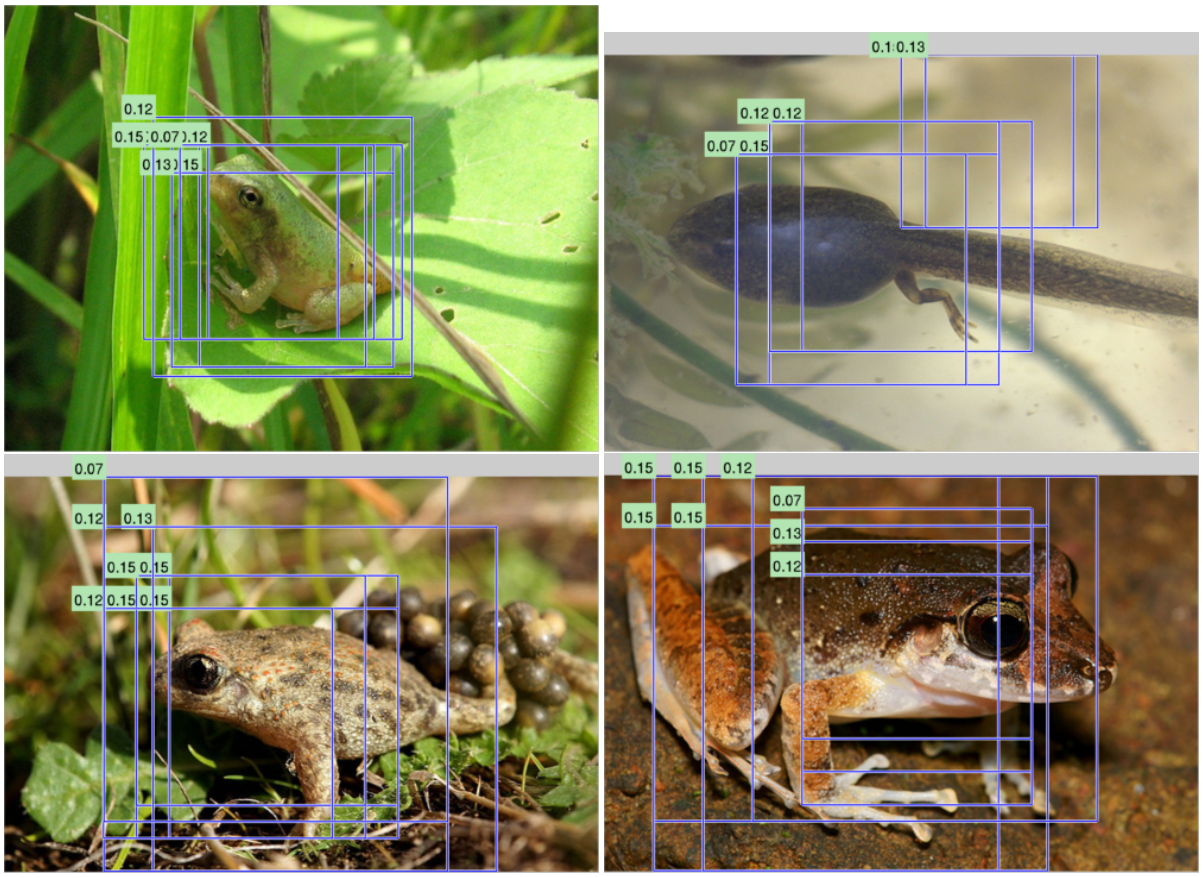

Figure 22: More examples of ILSVRC2014: n01644900(tailed frog)

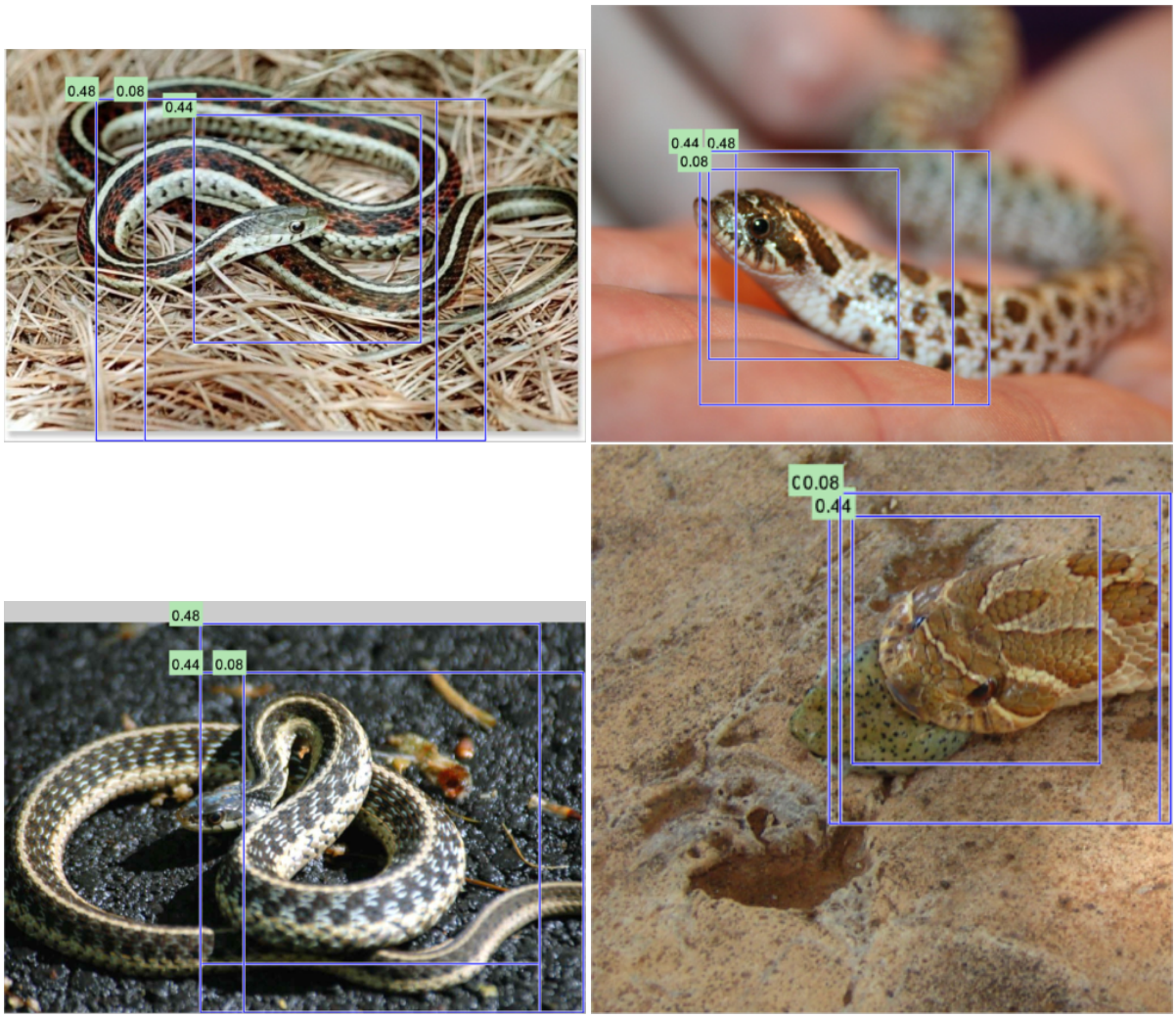

Figure 23: More examples of ILSVRC2014: n01729322(hognose snake).