

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 PASCAL 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.

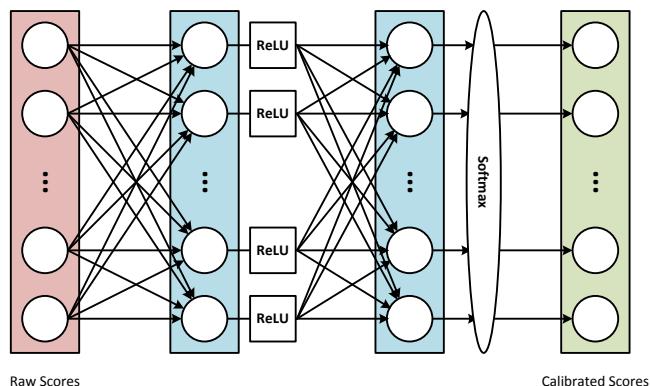


Figure 1: Three-Layer NN structure. The layers in purple are fully connected layers.

108	Model	jump.	phone.	playinstr.	read.	ridebike	ridehorse	run	takephoto	usecomp.	walk.	mAP	162
109	VGG16-LSSVM	85.5	64.5	95.2	71.6	93.0	95.9	82.5	69.0	90.7	62.0	81.0	163
110	VGG16-RRSVM	87.2	72.3	96.2	76.9	94.6	96.8	87.8	74.4	92.9	62.6	84.2	164
111	VGG19-LSSVM	85.5	64.5	95.4	72.4	93.0	96.0	83.4	70.9	90.7	62.1	81.4	165
112	VGG19-RRSVM	86.5	74.4	96.1	76.8	94.8	96.8	87.6	76.7	92.5	61.7	84.4	166
113	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	167

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.

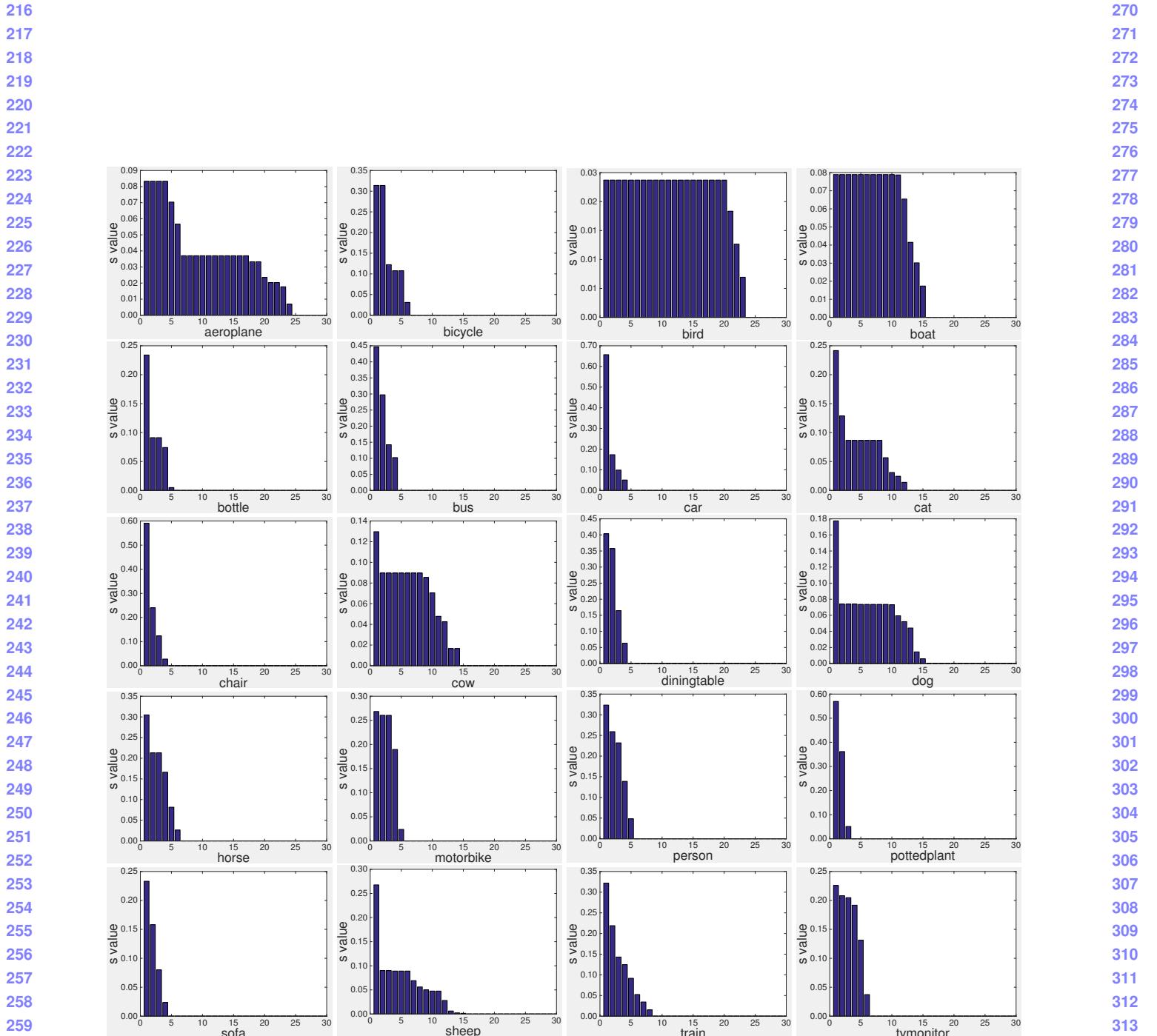
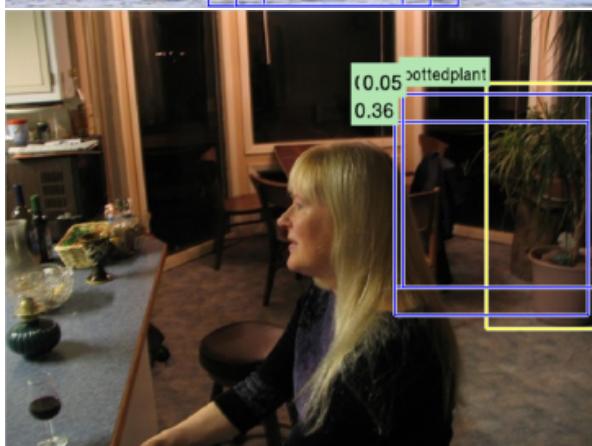
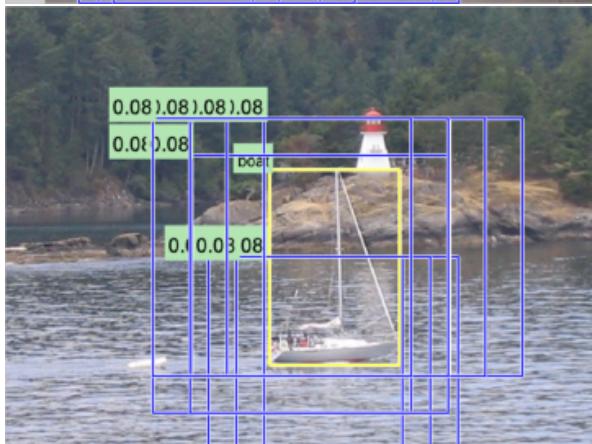
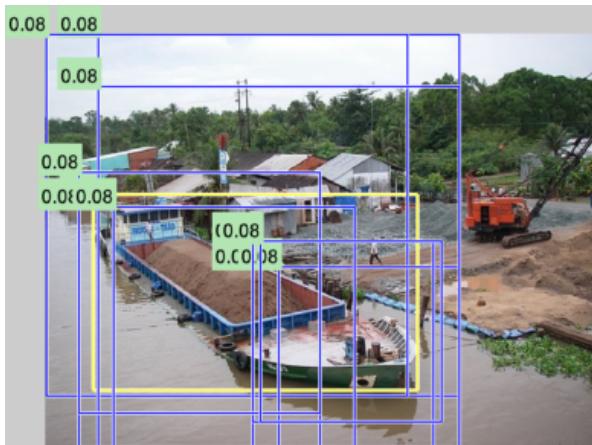
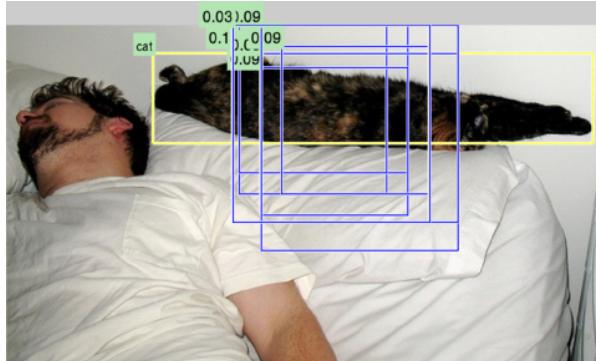
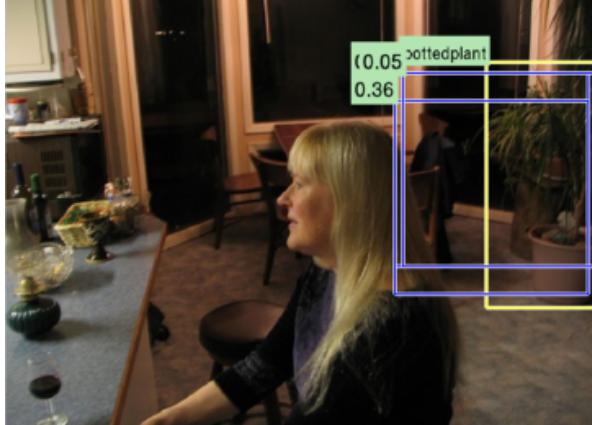
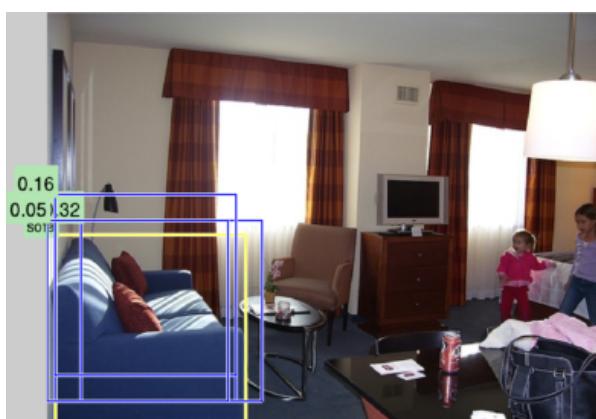


Figure 2: Distribution of region-score vectors for 20 categories in PASCAL VOC 2007 Image Classification dataset. Note that the effect number of regions for each category demonstrates the inadequacy of max-pooling or sum-pooling.

324
325
326
327
328
329
330

356



370

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.

371
372
373
374
375
376
377378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431

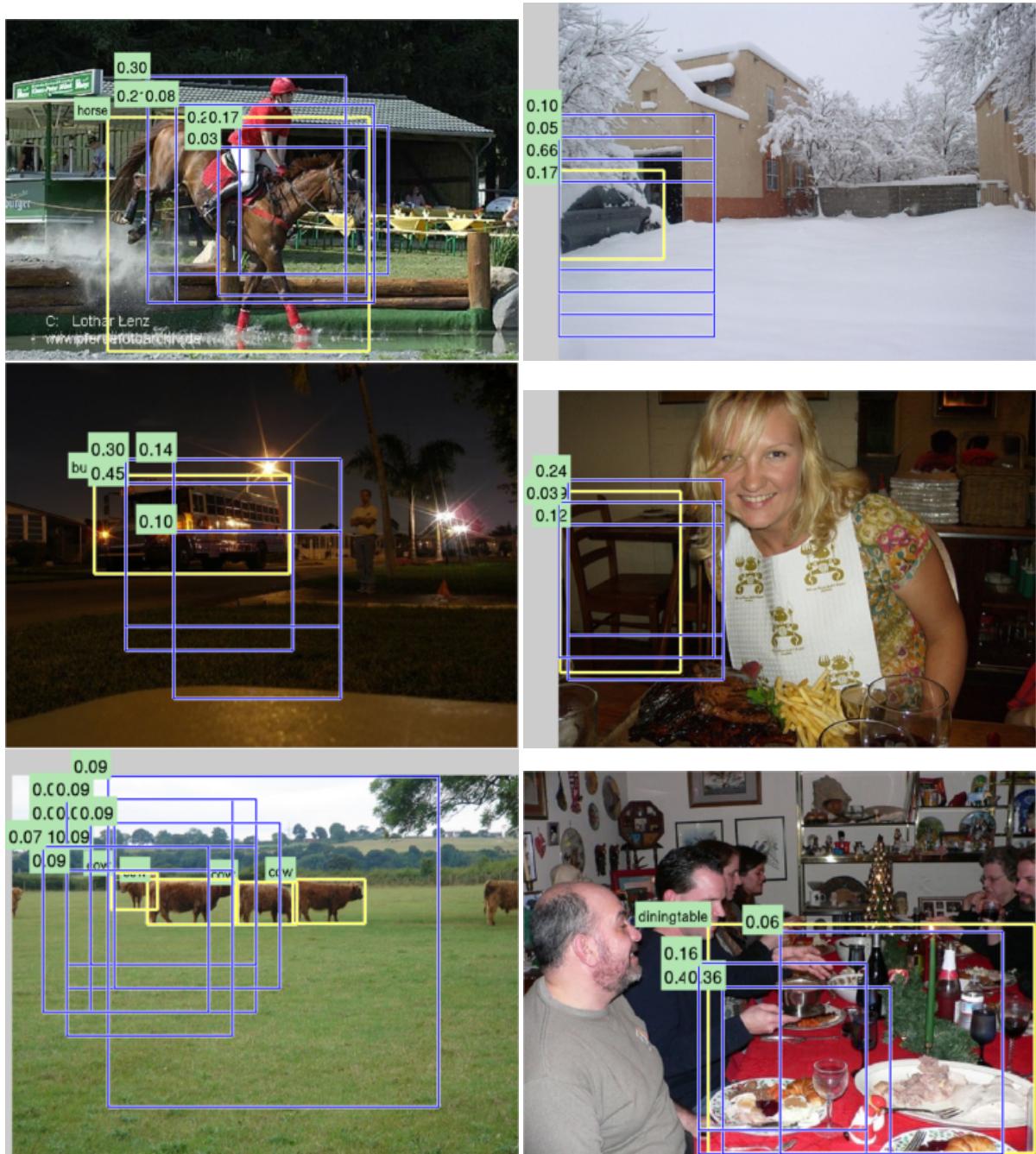
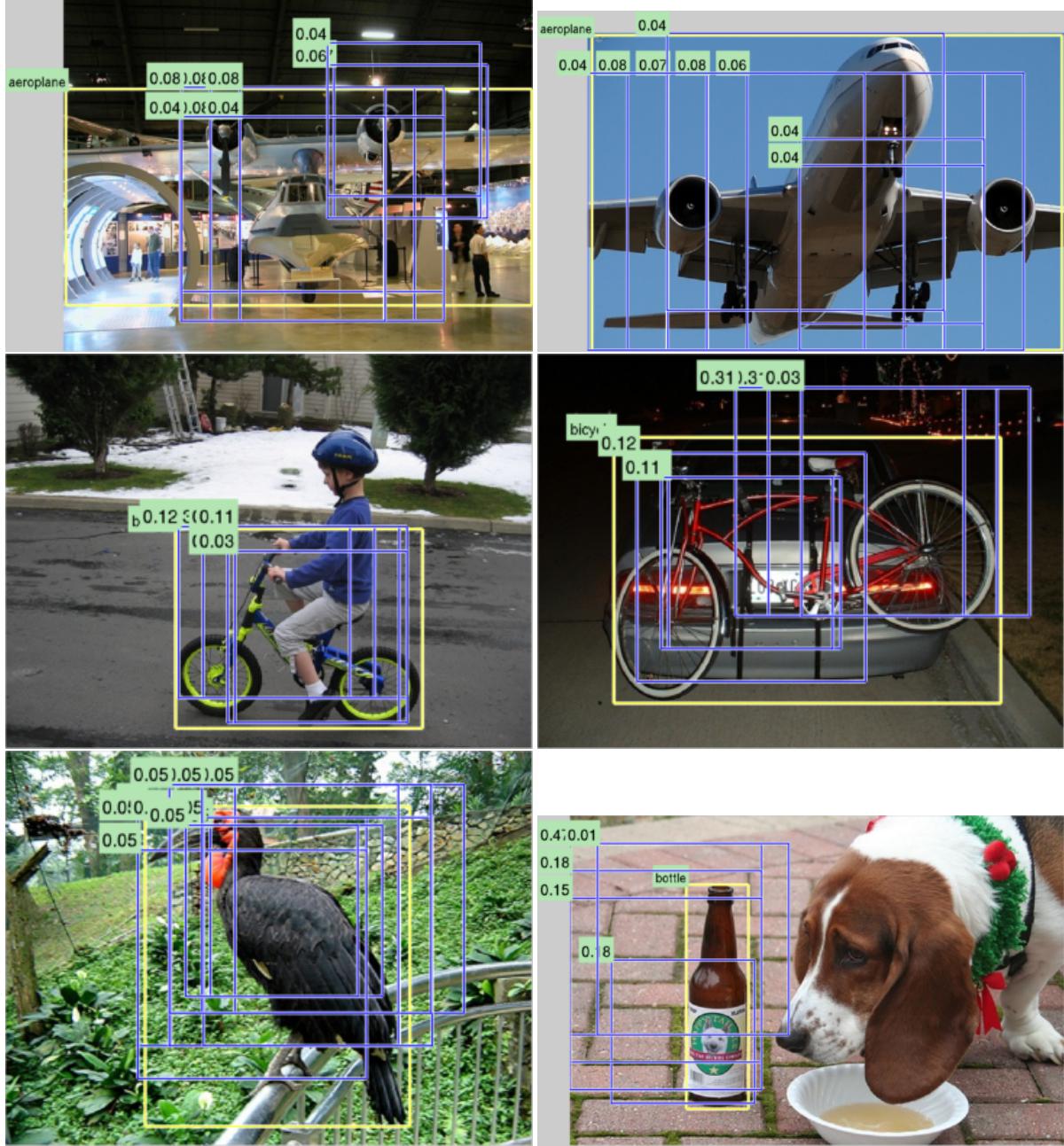


Figure 4: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific category. Best viewed on a digital device.

540
541
542
543
544

586 Figure 5: More images from the test set of PASCAL VOC 2007. The blue boxes with numbers correspond to the regions
587 and their weights used for classification. Yellow boxes with labels are the ground truth regions of interest for the specific
588 category. Best viewed on a digital device.
589
590
591
592
593

594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

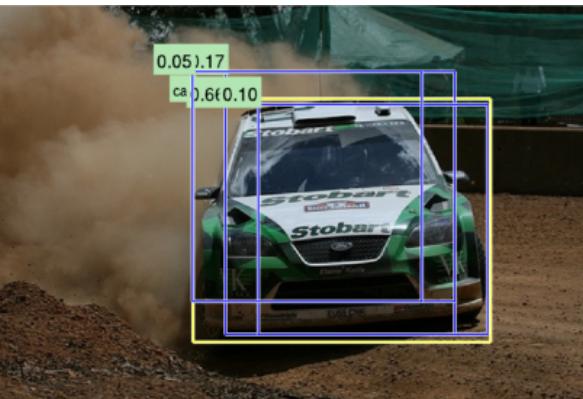
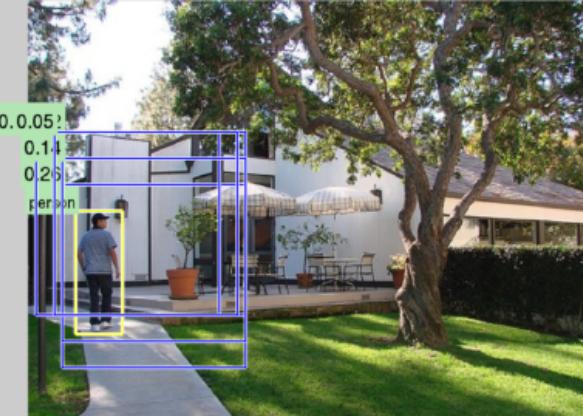
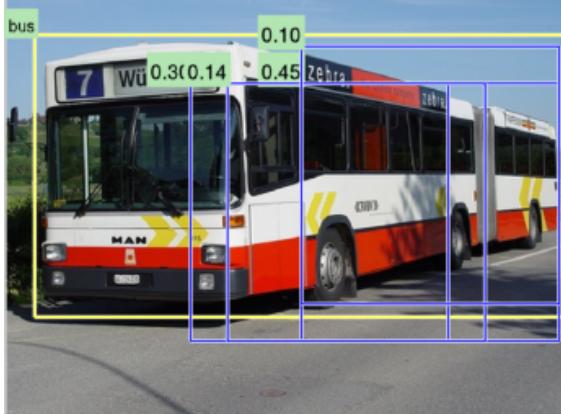
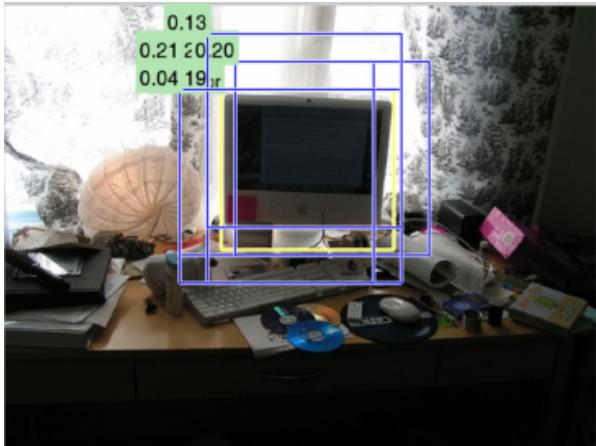
648
649
650
651

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.

698
699
700
701702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755

756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809

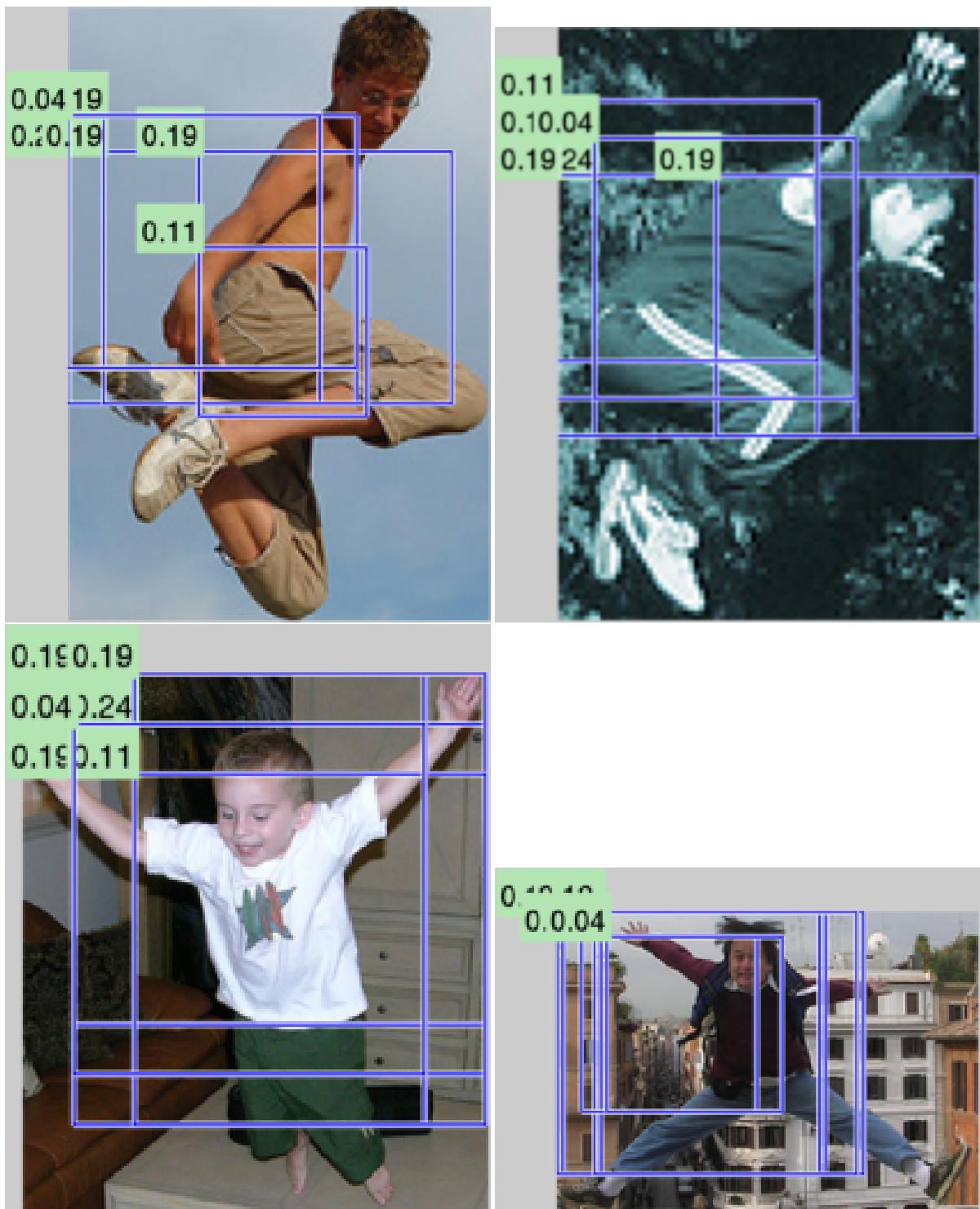
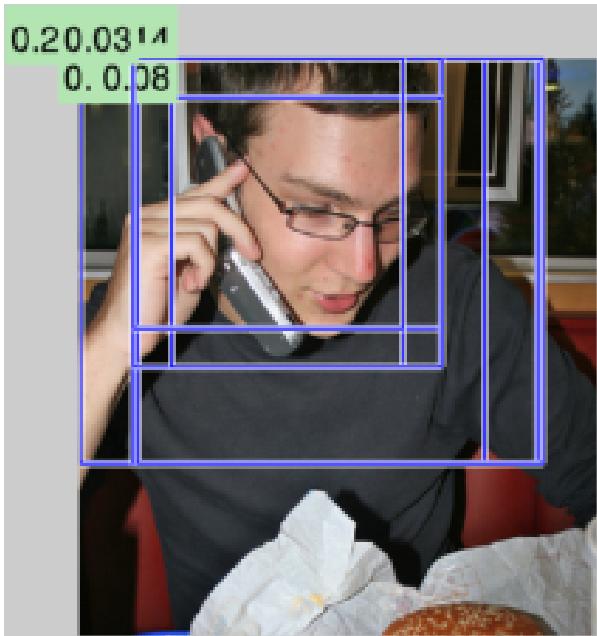
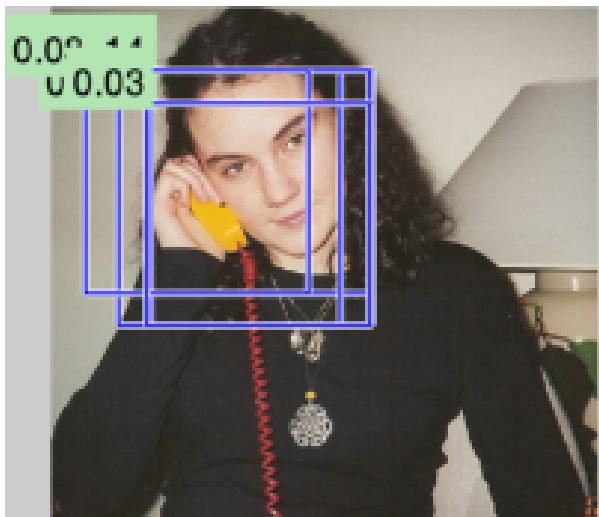
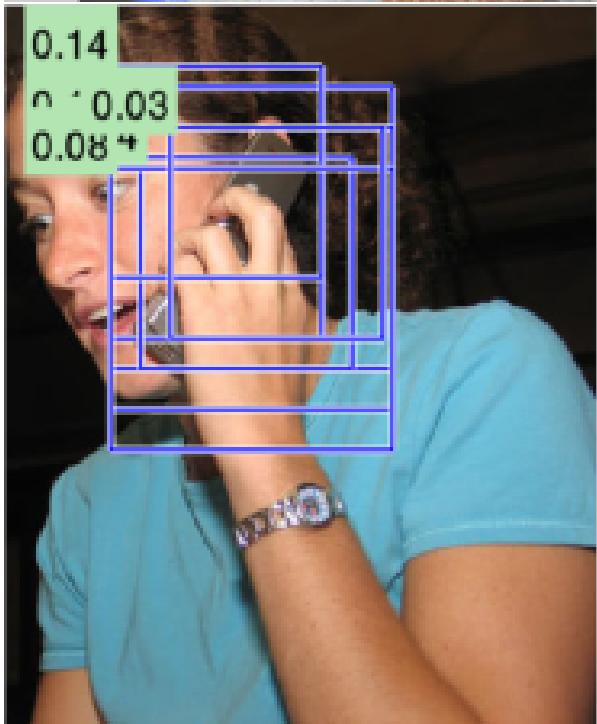
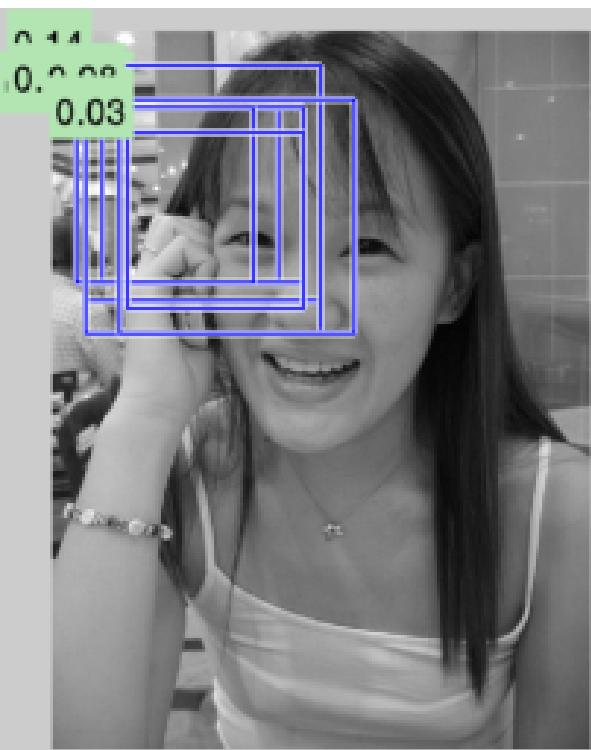


Figure 7: Examples of **jumping** in PASCAL VOC 2012 Action Classification.

810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863

864
865
866
867
868
869
870871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971Figure 8: Examples of **phoning** in PASCAL VOC 2012 Action Classification.

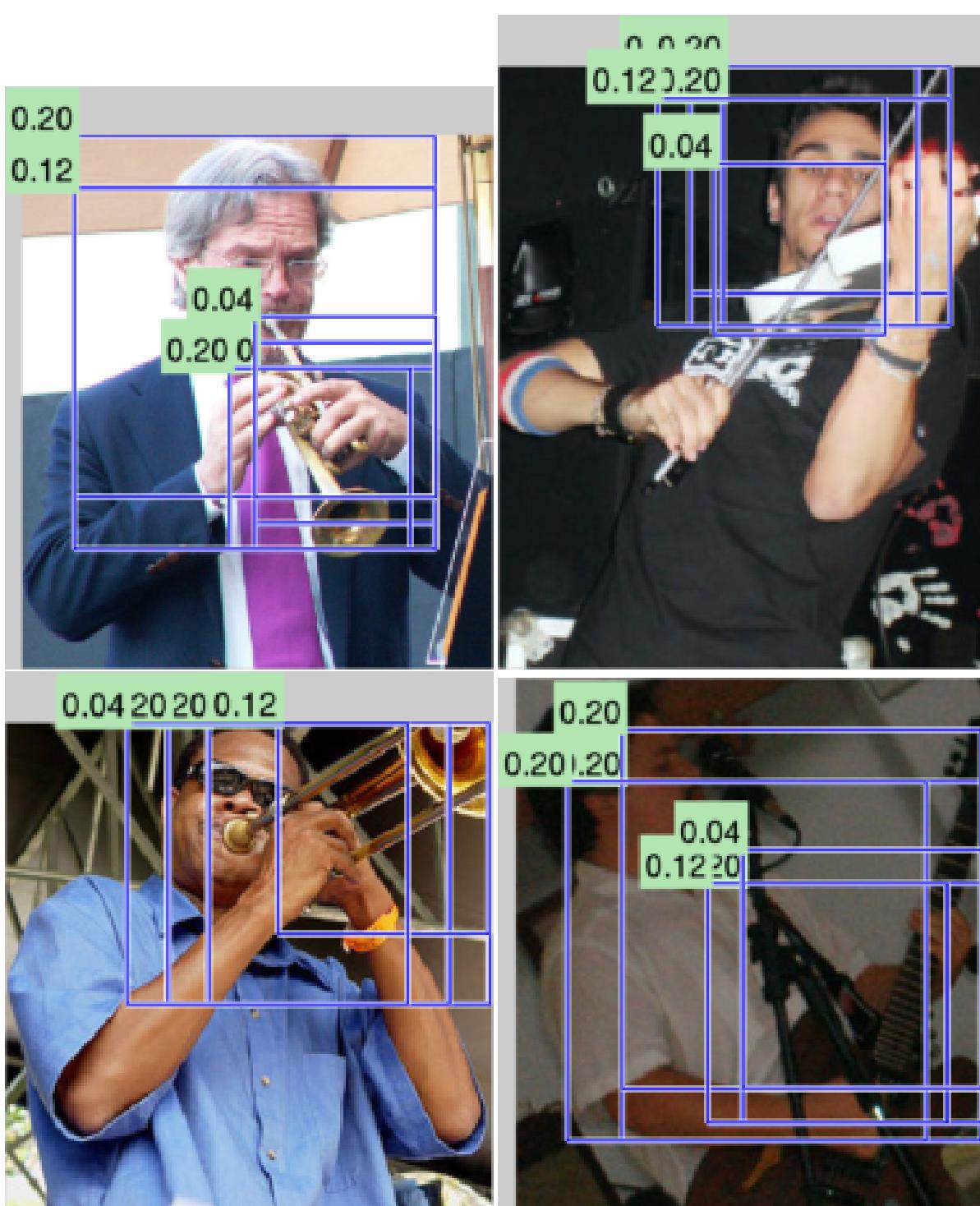


Figure 9: Examples of **playinginstrument** in PASCAL VOC 2012 Action Classification.

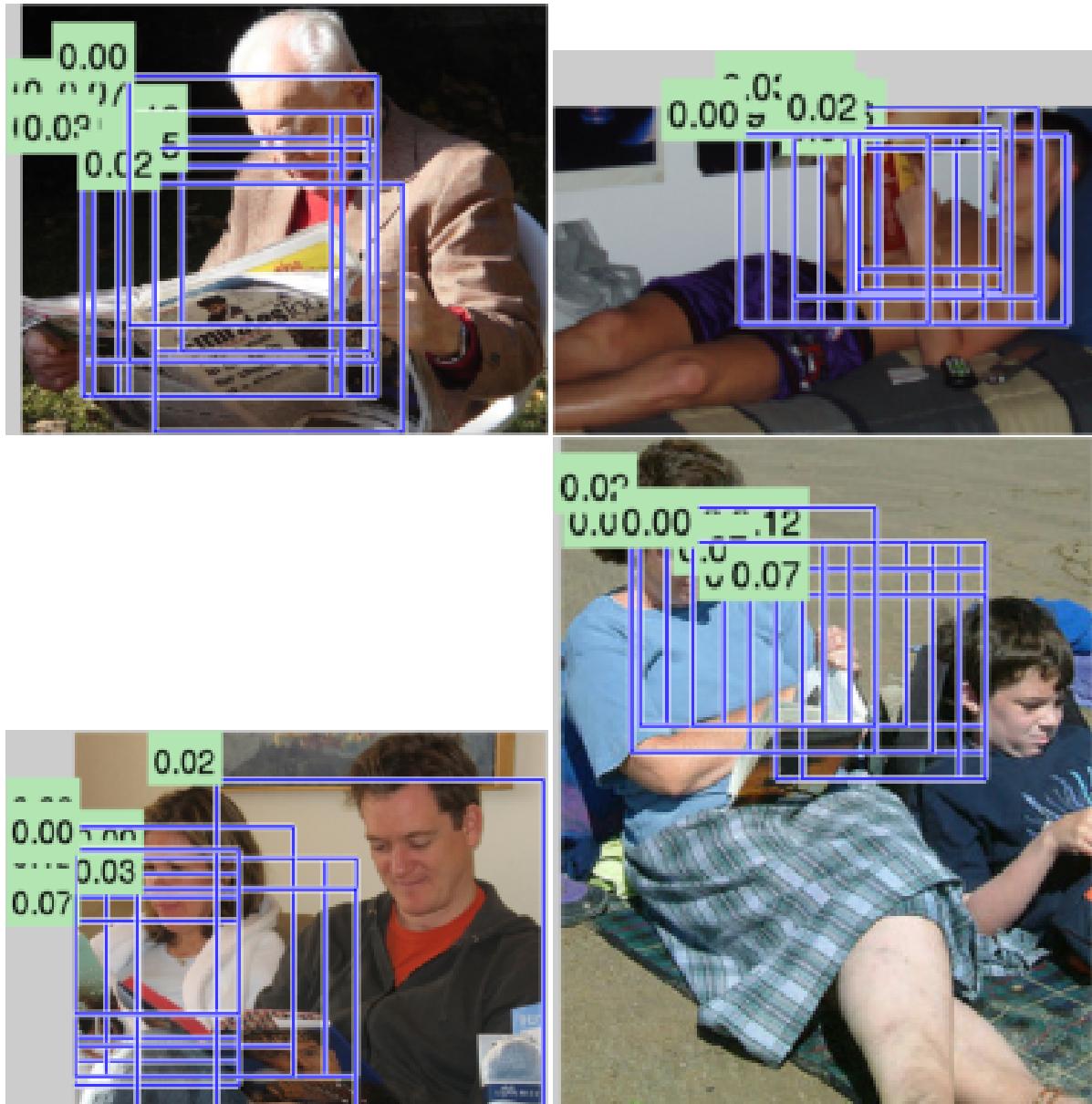


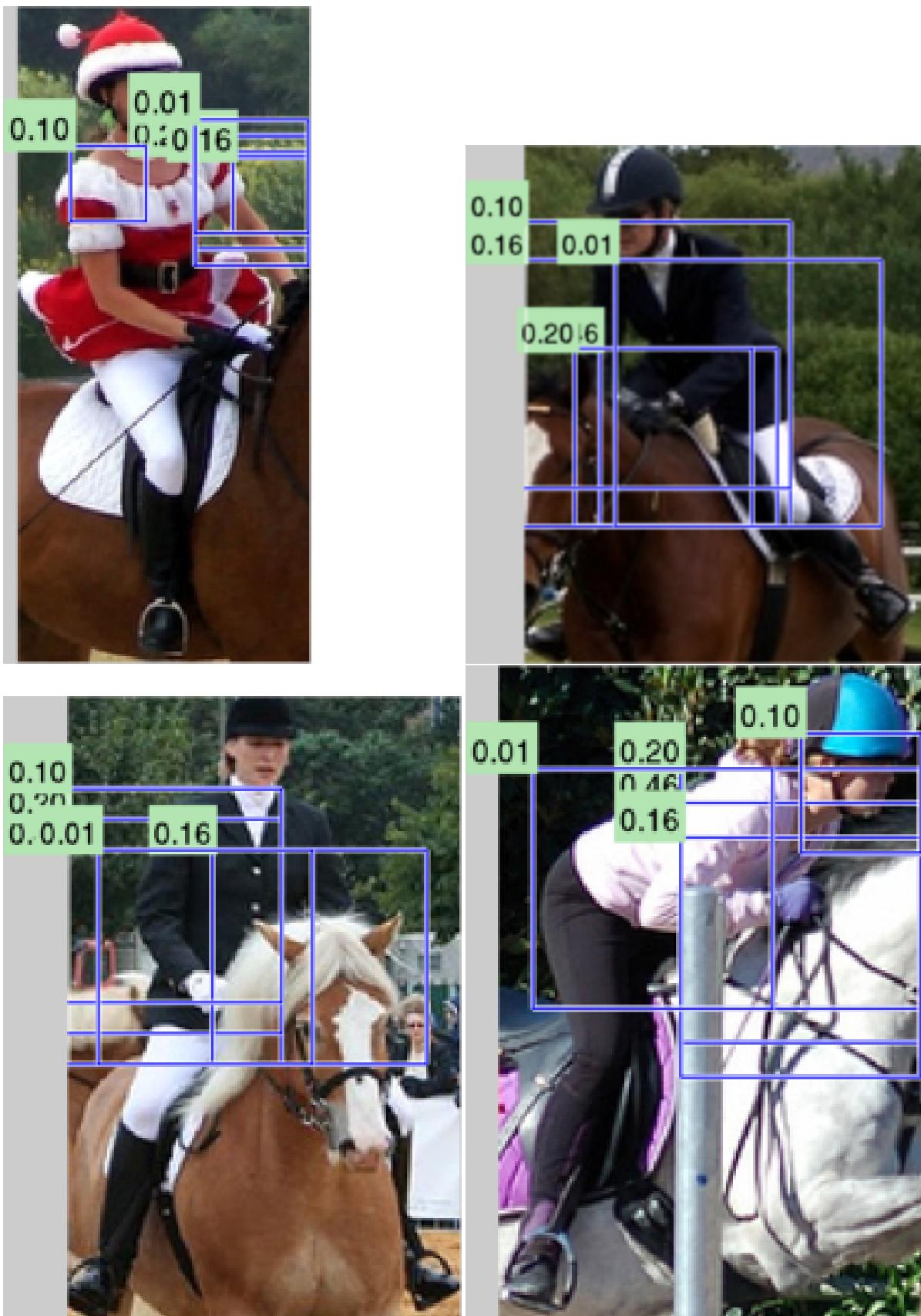
Figure 10: Examples of **reading** in PASCAL VOC 2012 Action Classification.

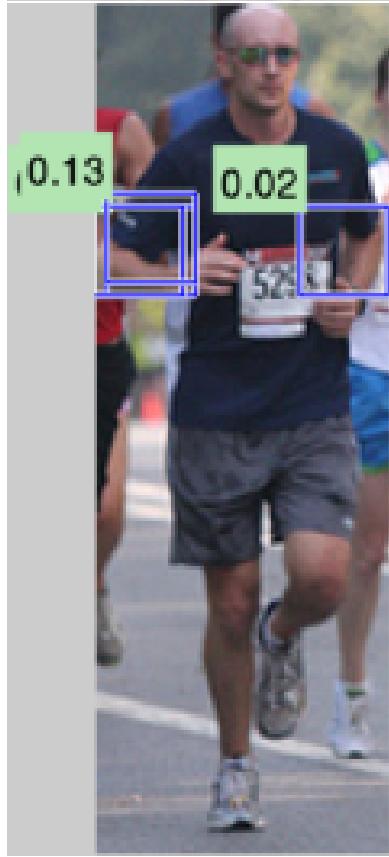
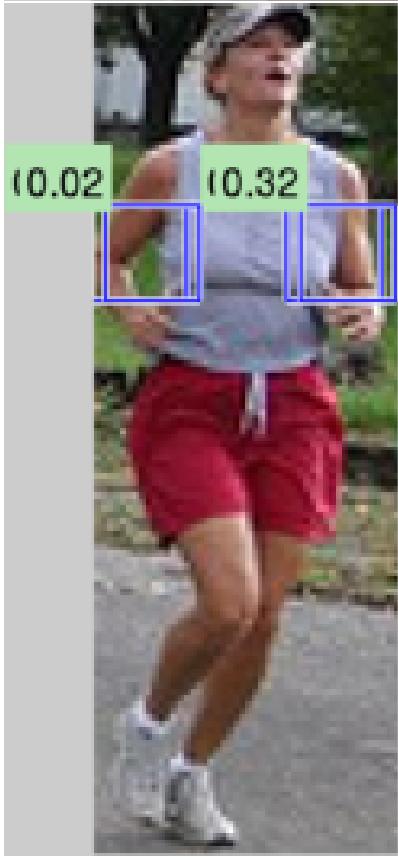
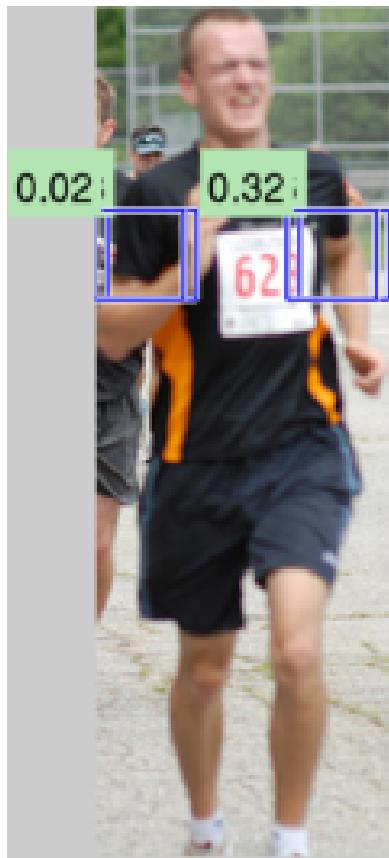
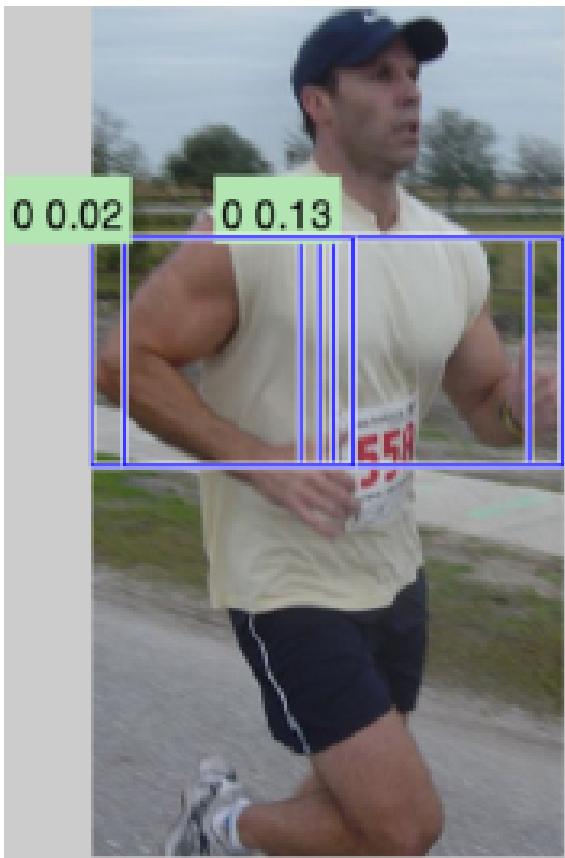
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241

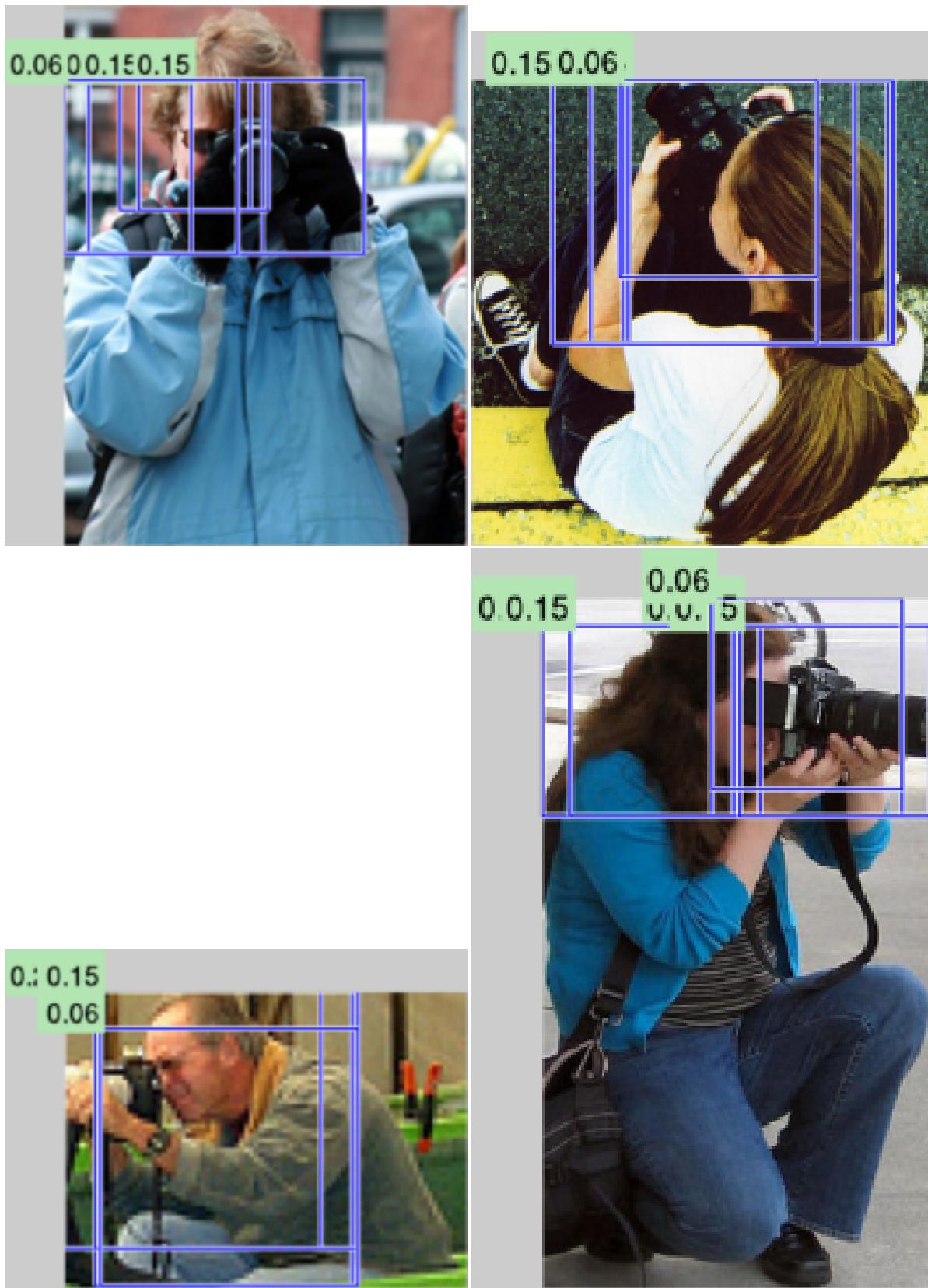
Figure 11: Examples of **ridingbike** in PASCAL VOC 2012 Action Classification.

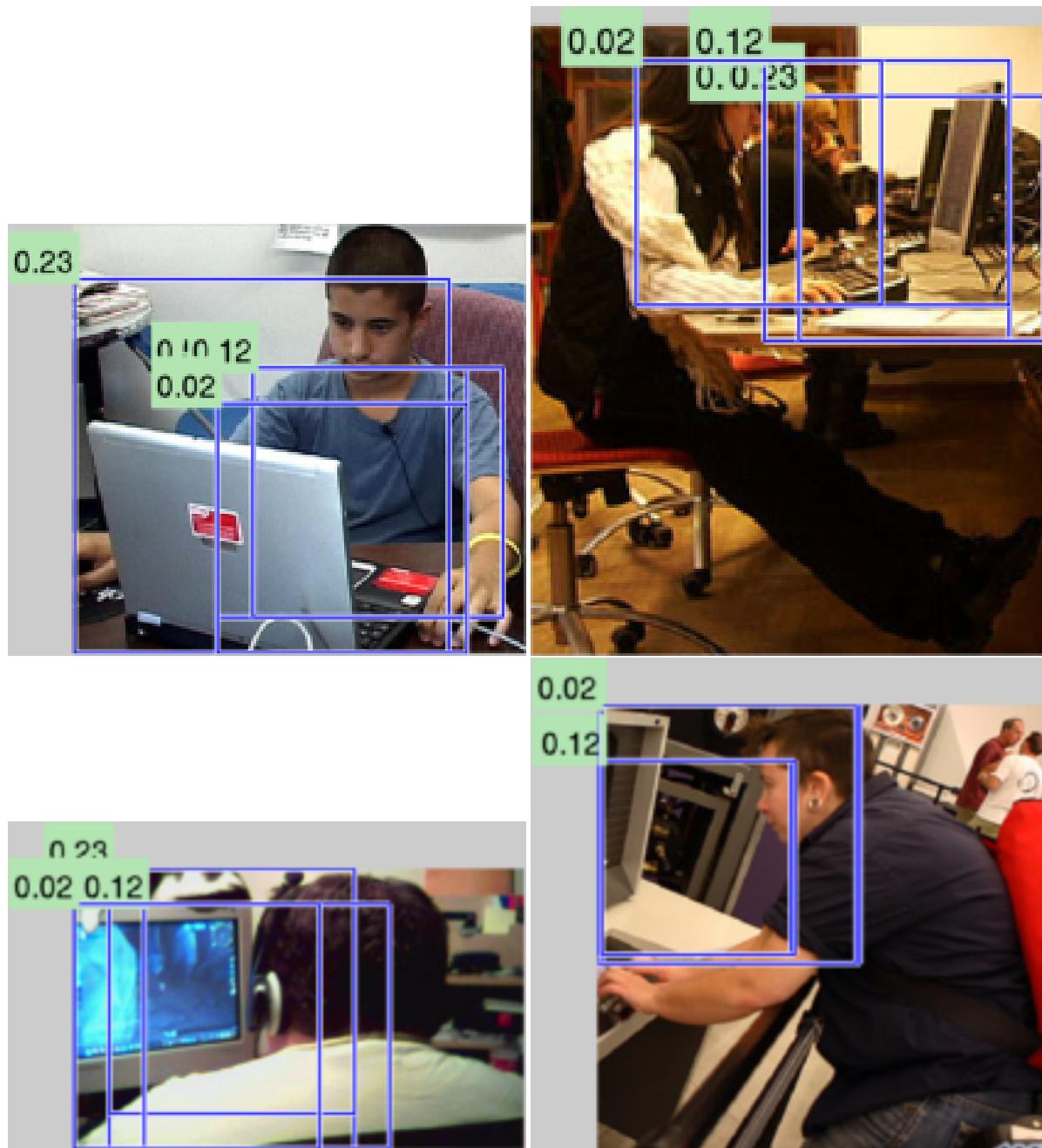
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295

1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349

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.

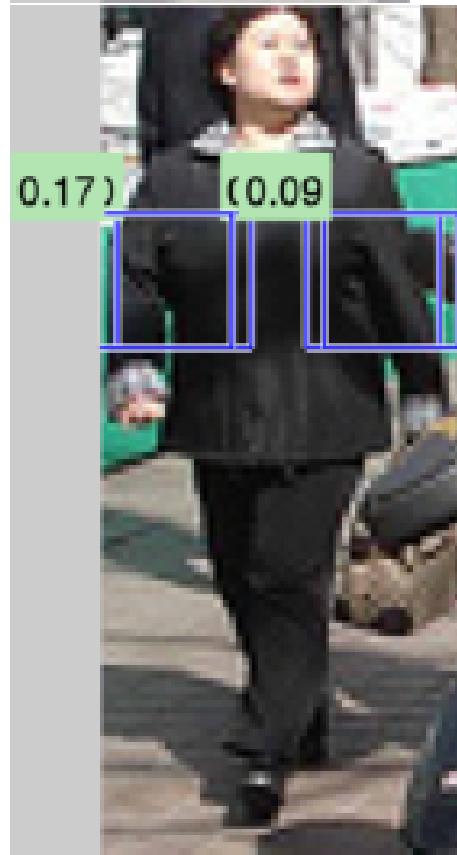
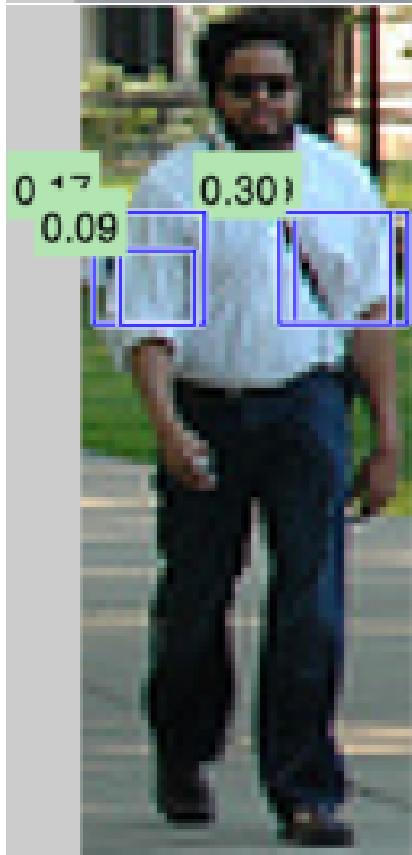
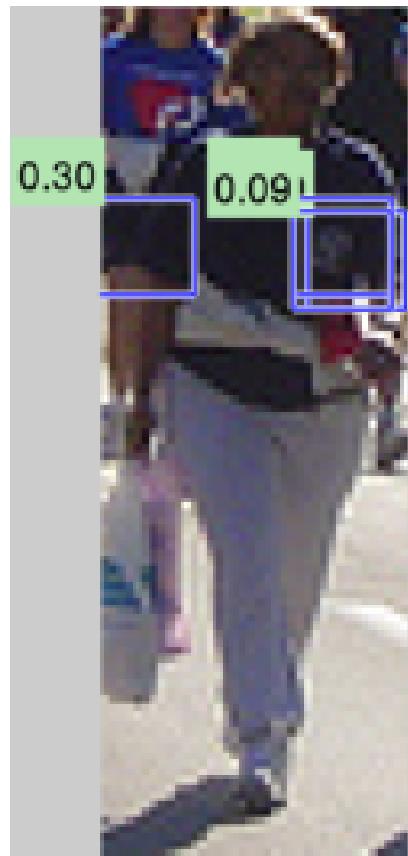
Figure 16: Examples of **walking** in PASCAL VOC 2012 Action Classification.



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.

CVPR
#2003

Figure 18 displays six images related to the category n02447721 (gong). The images are arranged in two rows of three. Each image features a blue bounding box and several green boxes containing numerical values, likely representing confidence scores or drop comparisons relative to a baseline.

- Top Row:**
 - Image 1:** A vertical gong with a black frame. Two green boxes at the top left and right corners contain the values 0.05 and 0.02 respectively.
 - Image 2:** A large, round gong mounted on a stand. A blue bounding box covers the entire gong, and a green box at the top left contains the value 0.05.
 - Image 3:** Two circular gongs placed side-by-side. A blue bounding box covers both, and green boxes at the top left and right contain the values 0.05 and 0.10 respectively.
- Bottom Row:**
 - Image 4:** A large, round gong mounted on a stand. A blue bounding box covers the entire gong, and green boxes at the top left and right contain the values 0.10 and 0.02 respectively.
 - Image 5:** A large, round gong mounted on a stand. A blue bounding box covers the entire gong, and green boxes at the top left and right contain the values 0.10 and 0.02 respectively.
 - Image 6:** A person playing a large gong with a mallet. A blue bounding box covers the gong, and green boxes at the top left and right contain the values 0.05 and 0.13 respectively.

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.

2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105

2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159

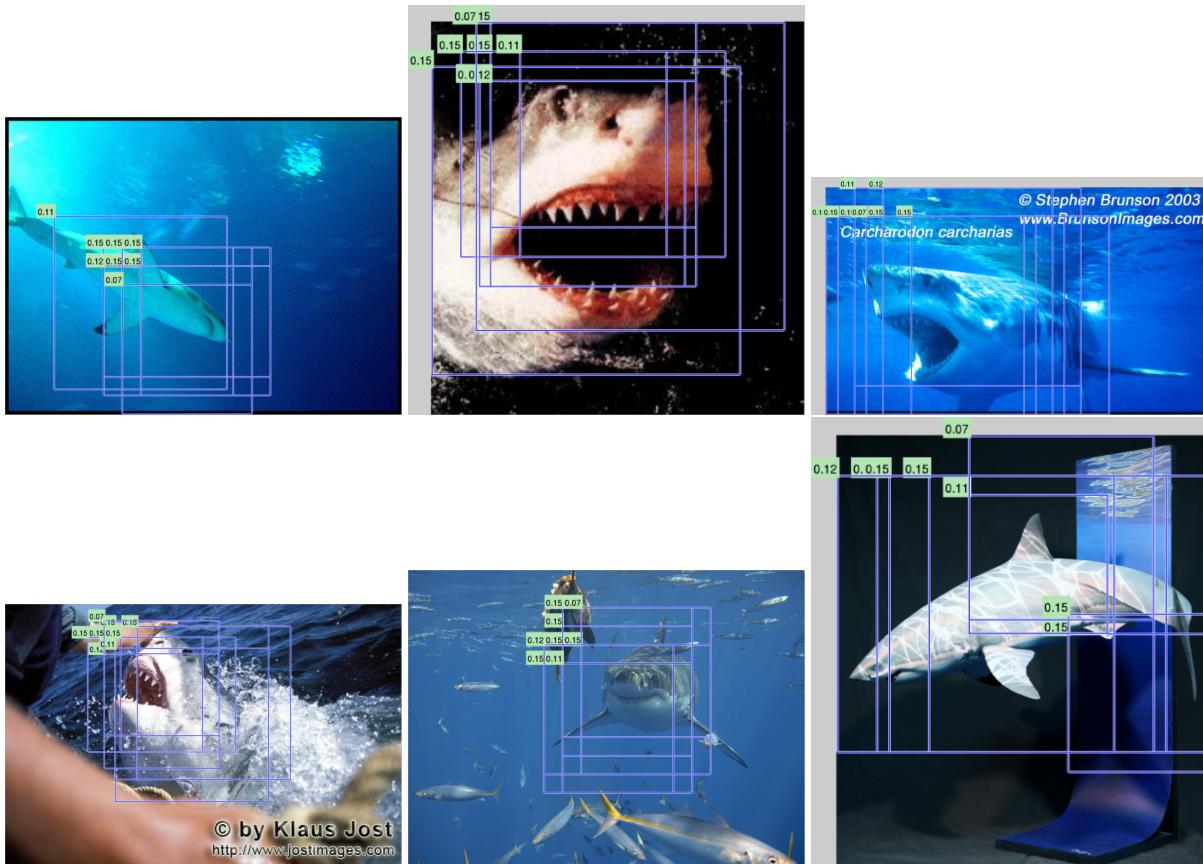


Figure 19: More examples of ILSVRC2014: n01484850 (white shark).

2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171

2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186

2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200

2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213

2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240

2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254

2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267

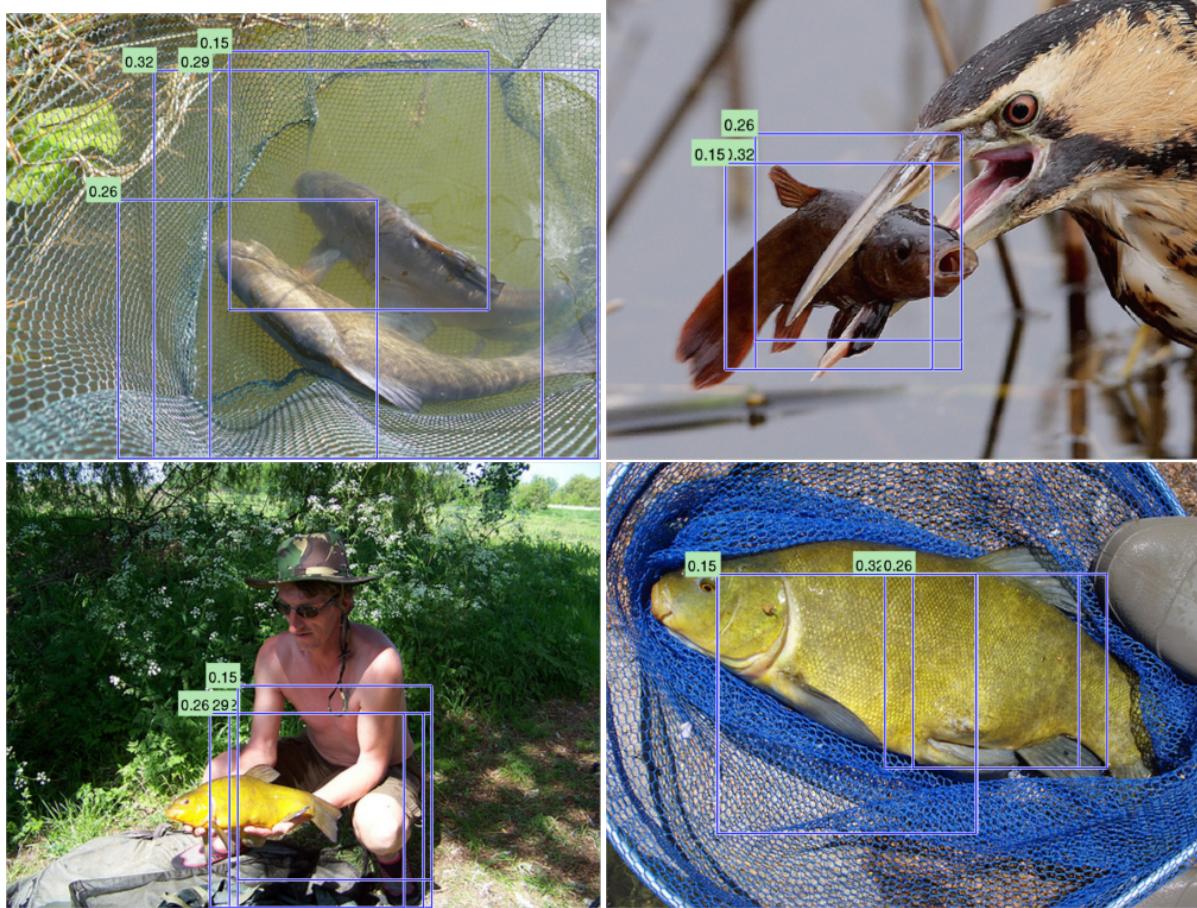


Figure 20: More examples of ILSVRC2014: n01440764(tench).

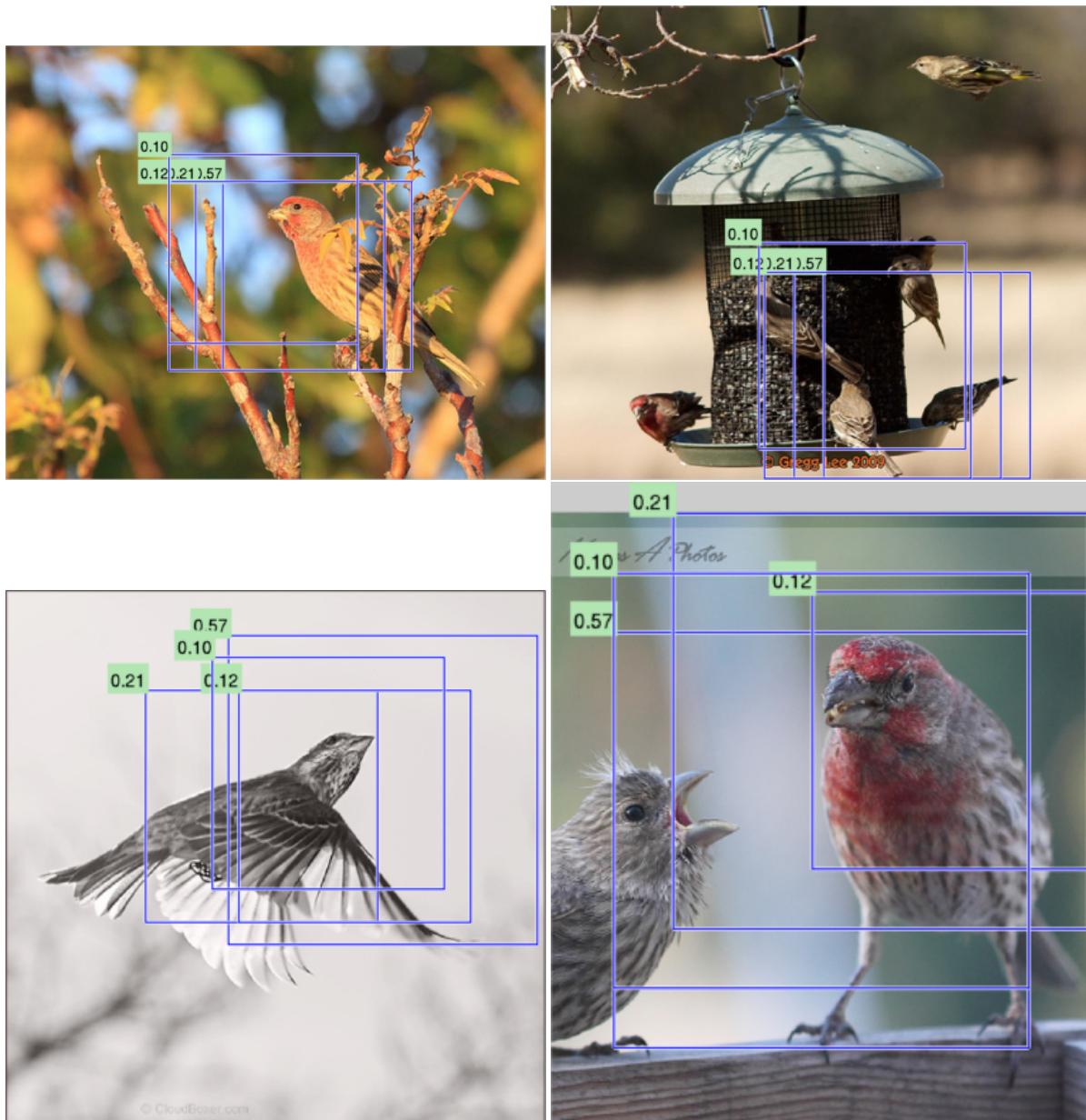


Figure 21: More examples of ILSVRC2014: n01532829(house finch).

2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321

2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375

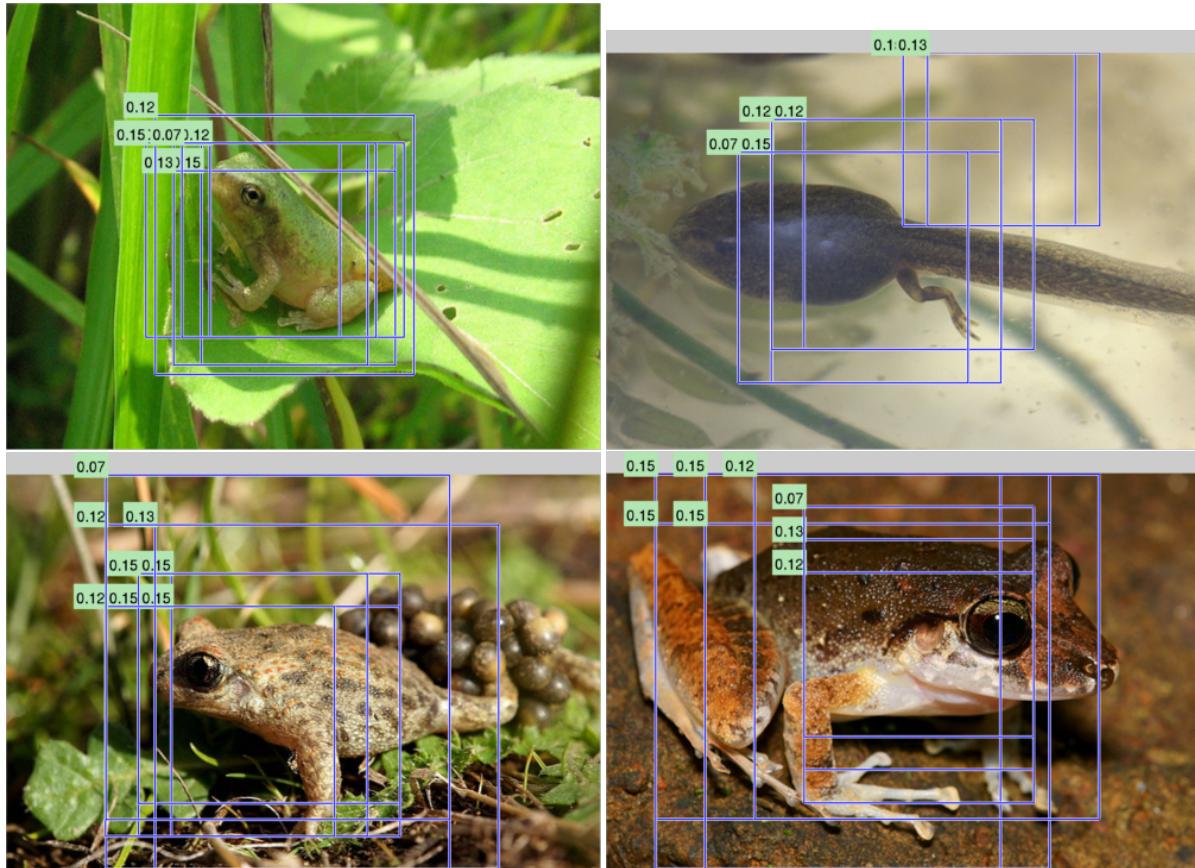


Figure 22: More examples of ILSVRC2014: n01644900(tailed frog)

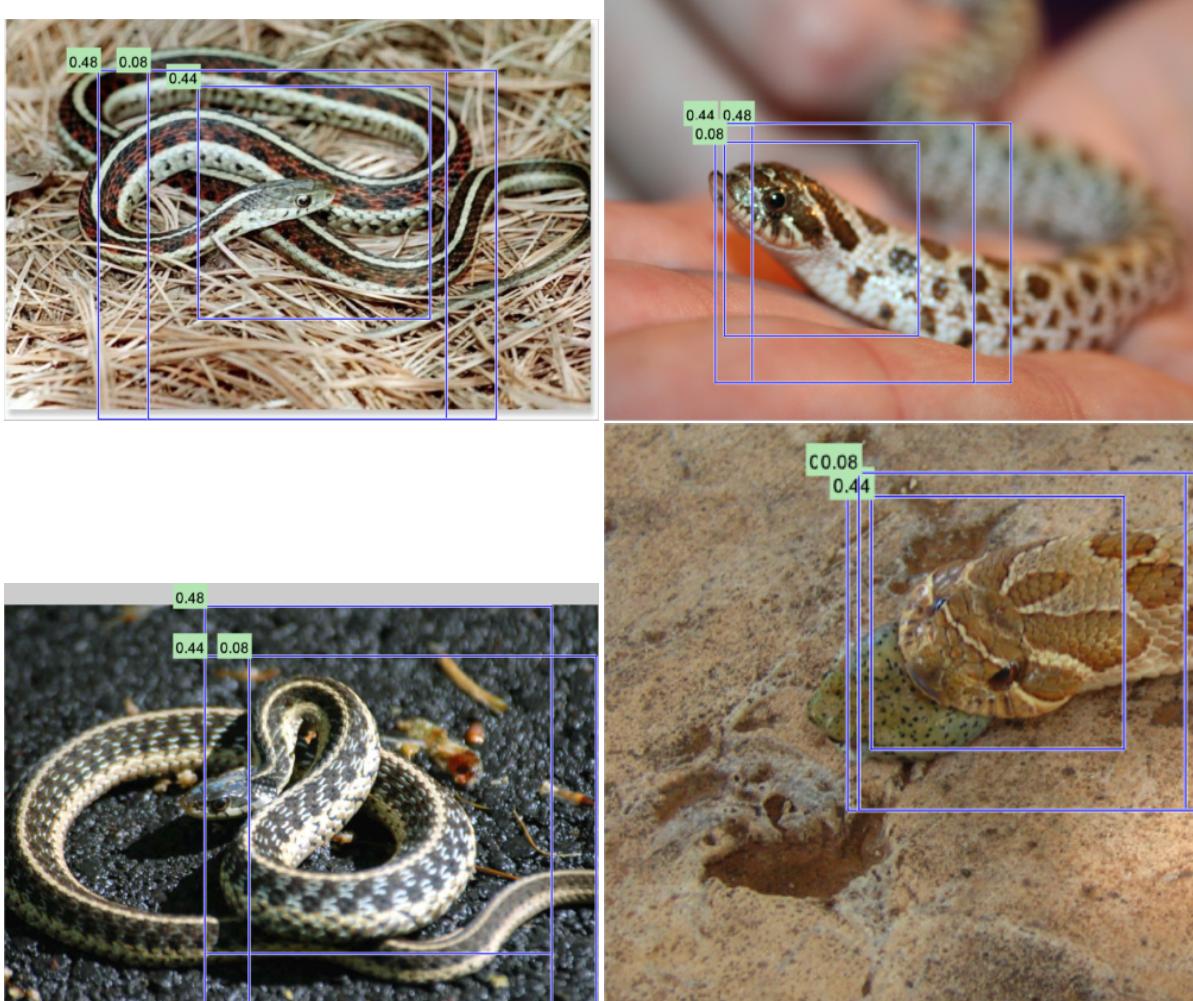


Figure 23: More examples of ILSVRC2014: n01729322(hognose snake).