

000
 001
 002
 003
 004 **Supplementary Material of**
 005 **Fisher Kernel for Deep Neural Activations**
 006
 007
 008
 009
 010
 011
 012
 013
 014
 015

016 Anonymous CVPR submission
 017
 018
 019
 020
 021
 022
 023
 024
 025
 026
 027
 028
 029
 030
 031
 032
 033
 034
 035
 036
 037
 038
 039
 040
 041
 042
 043
 044
 045
 046
 047
 048
 049
 050
 051
 052
 053

054
 055
 056
 057
 058
 059
 060
 061
 062
 063
 064
 065
 066
 067
 068
 069
 070
 071
 072
 073
 074
 075
 076
 077
 078
 079
 080
 081
 082
 083
 084
 085
 086
 087
 088
 089
 090
 091
 092
 093
 094
 095
 096
 097
 098
 099
 100
 101
 102
 103
 104
 105
 106
 107

1. Run-Time for Extracting Multi-scale Dense Activations

We measure a run-time for extracting multi-scale dense activations. We then compare the run-time of a naive way and that of the proposed way. The details of the each method are as follows.

Naive Extraction

The naive extraction is composed of two steps, preprocessing and feedforward. The preprocessing step includes cropping multi-scale dense patches and resizing each of them into the CNN input size (227×227 in [1]). For cropping and resizing, we use `imcrop` and `imresize` functions in MATLAB. In the feedforward step, each patch is fed into the CNN and an activation vector is obtained from a target layer (FC7). For fast extraction, we make multiple batches and feed each batch to the CNN, where each batch contains 256 patches.

Proposed Extraction

The proposed extraction is illustrated in Figure 2 in the submitted paper. It also includes preprocessing and feedforward steps. In the preprocessing step, we resize an image into multi-scale images, starting from 227×227 size as the smallest image scale. For resizing, we use `imresize` function in MATLAB. In the feedforward step, we feed each scaled image to the CNN and get multi-scale dense activation vectors.

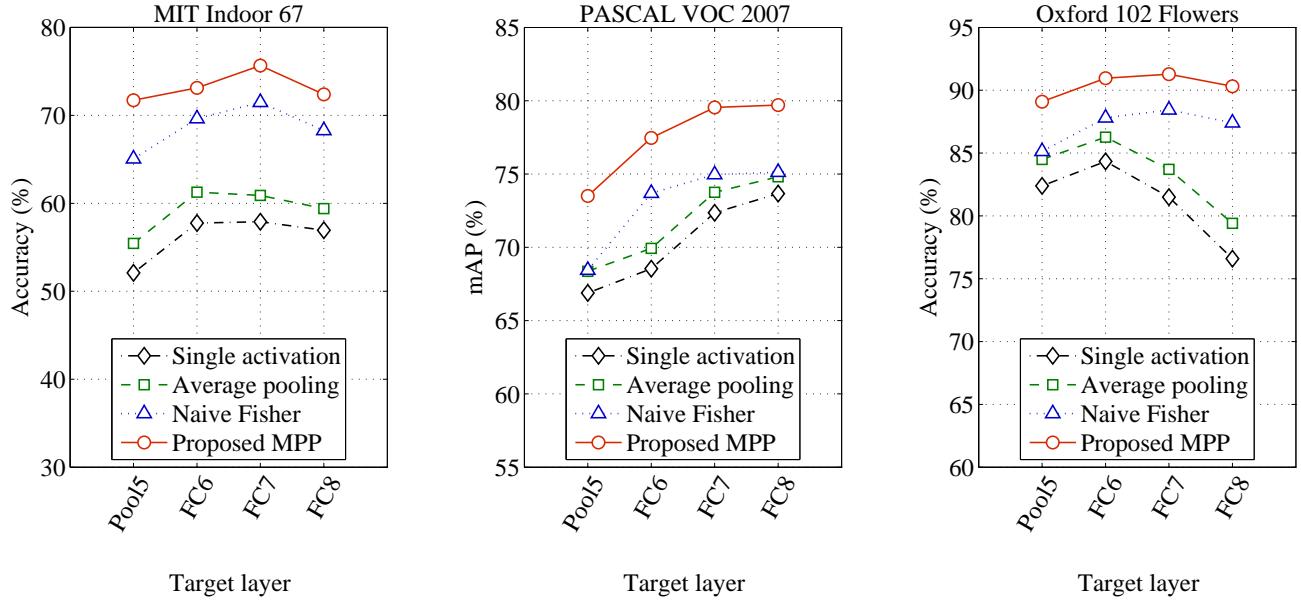
The preprocessing step is run on a CPU of 2.6GHz Intel Xeon. The extraction part is implemented with MatConvNet toolbox [2] and run on a GPU of GTX TITAN Black. The Caffe reference model [1] is used for the extraction part and the activations are obtained from FC7 layer of the model. We measure the average run-time with 100 random images from PASCAL VOC 2007. The result is shown in Table 1. With the proposed method, *a substantial amount of rich activations are obtained with reasonable time*.

Combination of image scales	1~4	1~5	1~6	1~7
Number of activations	270	754	1,910	4,410
Performance (%) on MIT Indoor 67	68.58	71.34	74.33	75.67
Performance (%) on PASCAL VOC 2007	78.06	79.16	79.65	79.54
Performance (%) on Oxford 102 Flowers	87.88	89.79	90.86	91.28
Preprocessing time (sec) for the naive extraction	1.008	2.942	6.478	16.28
Preprocessing time (sec) for the proposed extraction	0.0457	0.0741	0.1016	0.1872
Feedforward time (sec) for the naive extraction	0.6934	1.998	4.932	11.36
Feedforward time (sec) for the proposed extraction	0.0312	0.0524	0.1405	0.2763
Total time (sec) for the naive extraction	1.702	4.941	11.41	27.64
Total time (sec) for the proposed extraction	0.0769	0.1265	0.2420	0.4635
Speedup (times)	$\times 22.13$	$\times 39.06$	$\times 47.15$	$\times 59.63$

Table 1. Average run-time for extracting multi-scale dense activations per image.

108 **2. Performances with Different Target Layers** 162
109

110 We show performances of classification with different target layers. We use Caffe reference model [1] and seven scales
111 for image pyramid. The PCA dimension is 128 and the number of Gaussians is 256. For average pooling method, ten images
112 are augmented from one image.

132 Figure 1. Performances with different target layers. 186
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

216

3. Performances with Different Fisher Kernel Parameters

217

Every experiment here uses activations from FC7 of the Caffe reference model [1] and seven scales for image pyramid. We decided to use PCA dimension of 128 and 256 Gaussians for efficient experiments in the submitted paper, because the two values (128, 256) show reasonably good performance in all benchmarks, as shown in Fig. 2 and Fig. 3.

221

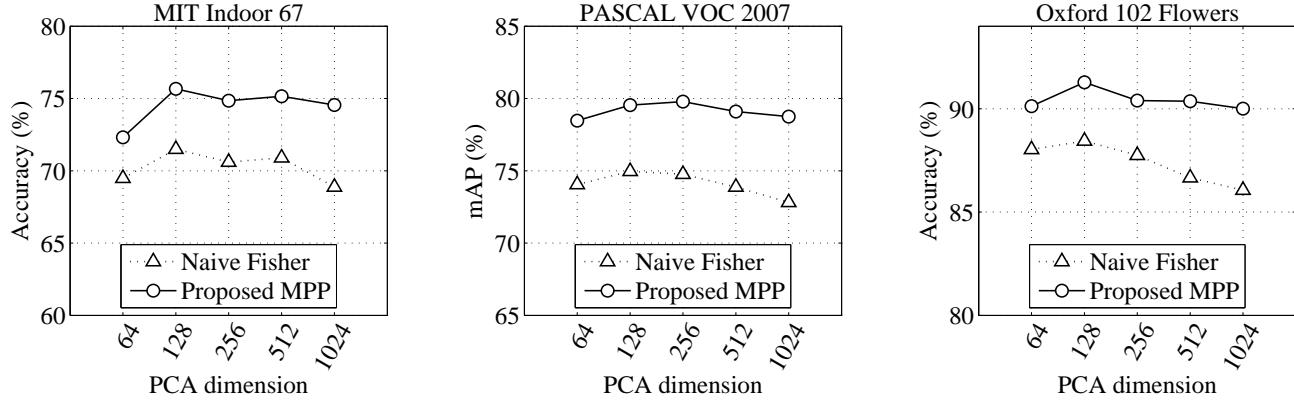


Figure 2. Performances with different PCA dimensions.

235

236

237

238

239

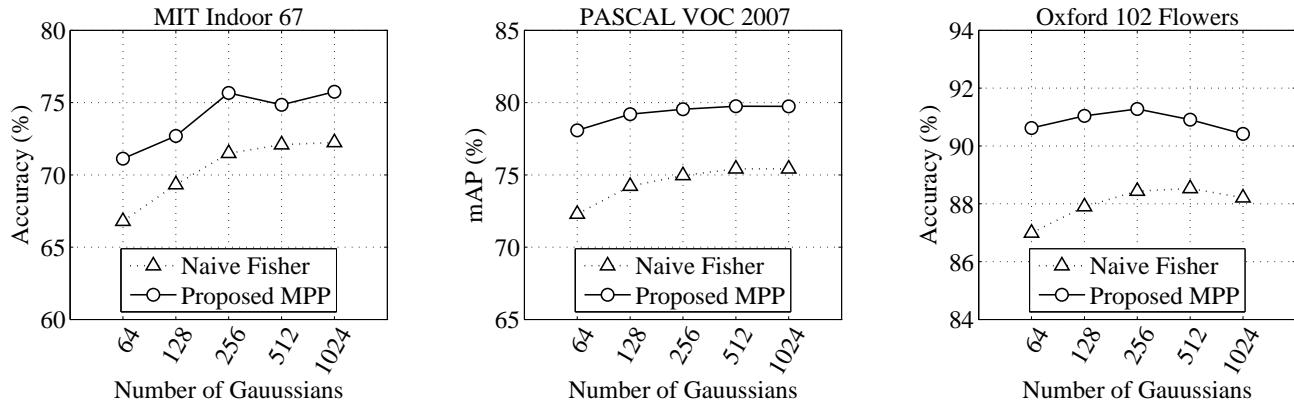


Figure 3. Performances with different number of Gaussians.

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324	4. More Examples of Weakly-Supervised Object Confidence Maps	378
325		379
326	We show more examples of object confidence map from PASCAL VOC 2007. How we obtain the object confidence maps	380
327	is described in Sec. 4.5 in the submitted paper. Note that <i>the following confidence maps are obtained by the supervision only</i>	381
328	<i>with the class-level labels, without using object bounding box annotations.</i>	382
329	In Fig. 4 and Fig. 5, we show examples of a single image having different multiple classes. In Fig. 6 ~ Fig. 25, we	383
330	show good and bad examples of 20 different classes. Every image in the examples are test images, not training images. The	384
331	examples start in the next page.	385
332		386
333		387
334		388
335		389
336		390
337		391
338		392
339		393
340		394
341		395
342		396
343		397
344		398
345		399
346		400
347		401
348		402
349		403
350		404
351		405
352		406
353		407
354		408
355		409
356		410
357		411
358		412
359		413
360		414
361		415
362		416
363		417
364		418
365		419
366		420
367		421
368		422
369		423
370		424
371		425
372		426
373		427
374		428
375		429
376		430
377		431

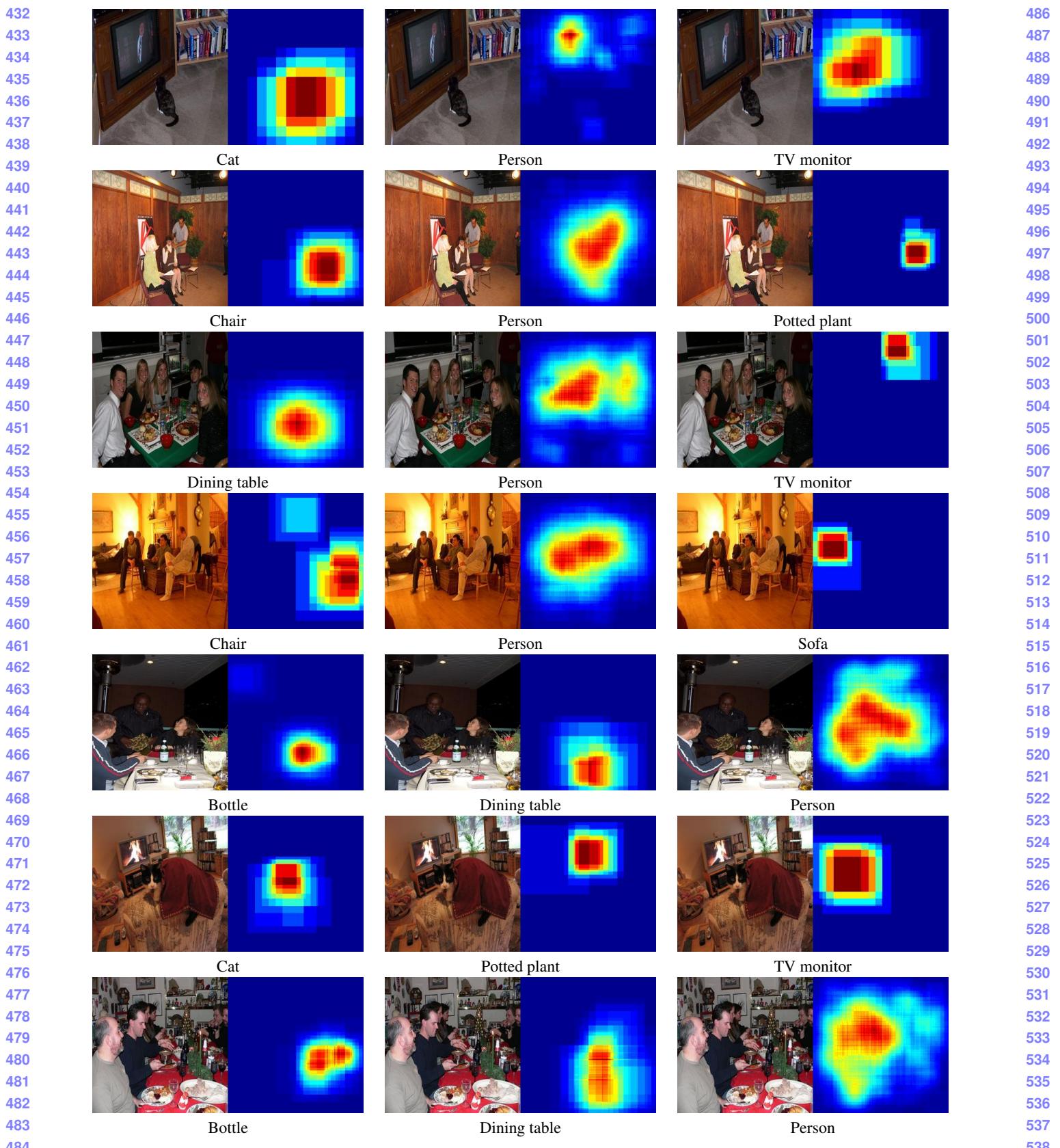


Figure 4. Examples of multi-class object confidence map of indoor scenes.

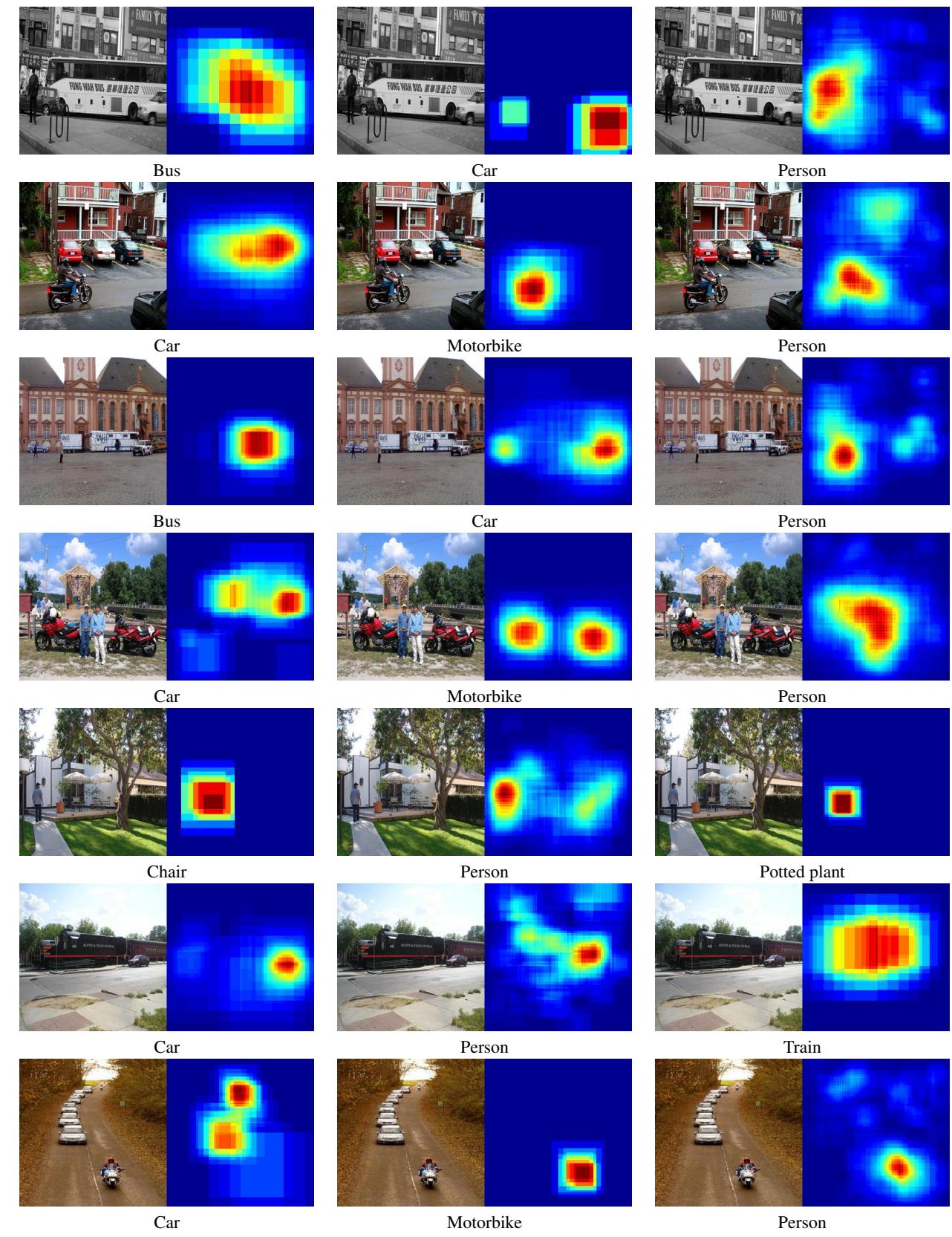
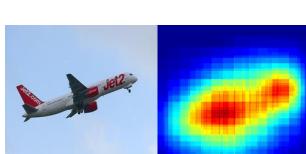
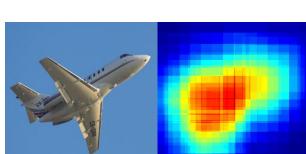
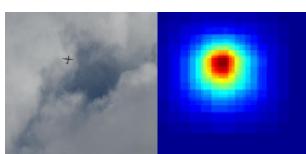
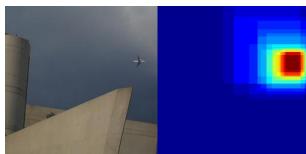


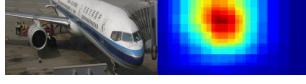
Figure 5. Examples of multi-class object confidence map of outdoor scenes.

648
649
650
651
652
653
654
655
656
657
658
659
660



672
673
674
675
676
677

678
679
680
681
682
683



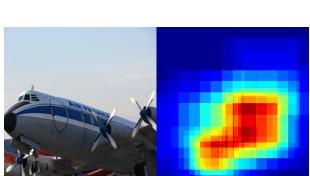
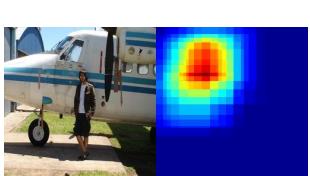
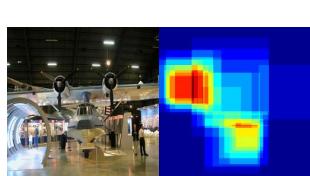
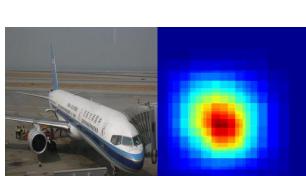
684
685

(a) True examples of the “aeroplane” class.

686
687
688
689
690
691
692

Figure 6. Examples of object confidence map of “aeroplane” class in the PASCAL VOC 2007 test image set.

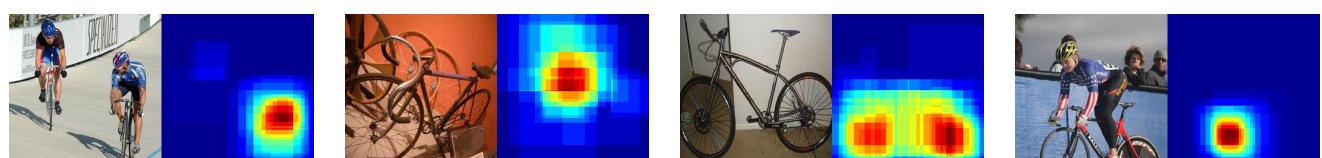
693
694
695
696
697
698
699
700
701



702
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

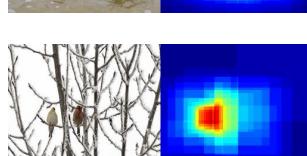
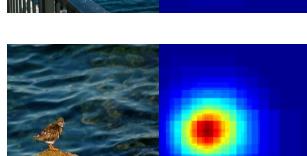
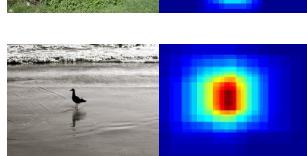
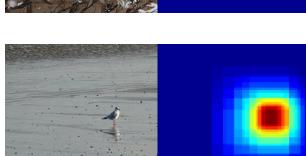
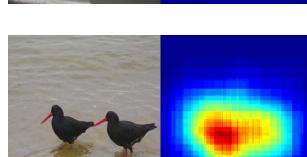
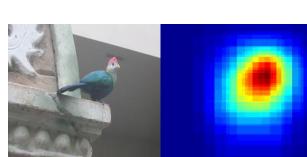
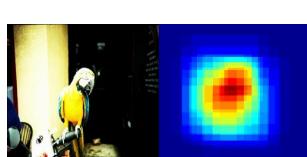
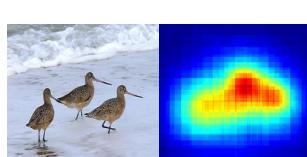
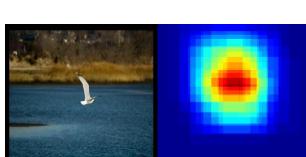
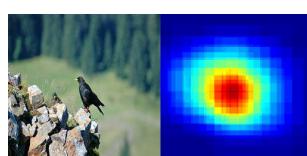
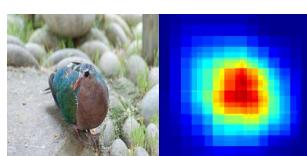
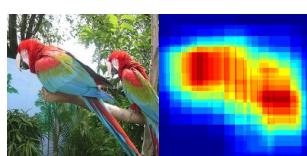
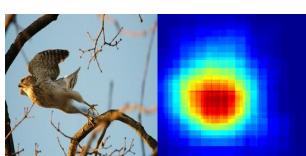
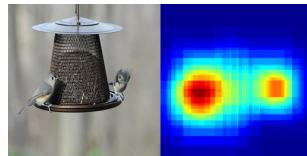
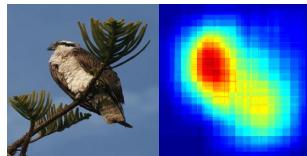
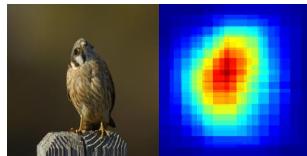
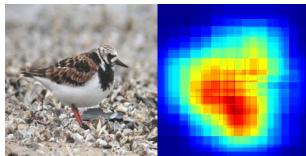


(a) True examples of the “bicycle” class.

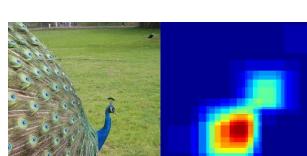
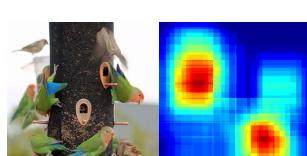
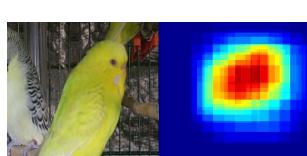
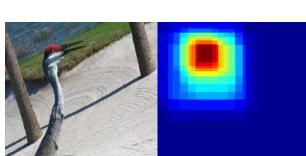


(b) False examples of the “bicycle” class.

Figure 7. Examples of object confidence map of “bicycle” class in the PASCAL VOC 2007 test image set.

864
865
866
867
868
869
870

(a) True examples of the “bird” class.



(b) False examples of the “bird” class.

Figure 8. Examples of object confidence map of “bird” class in the PASCAL VOC 2007 test image set.

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
971

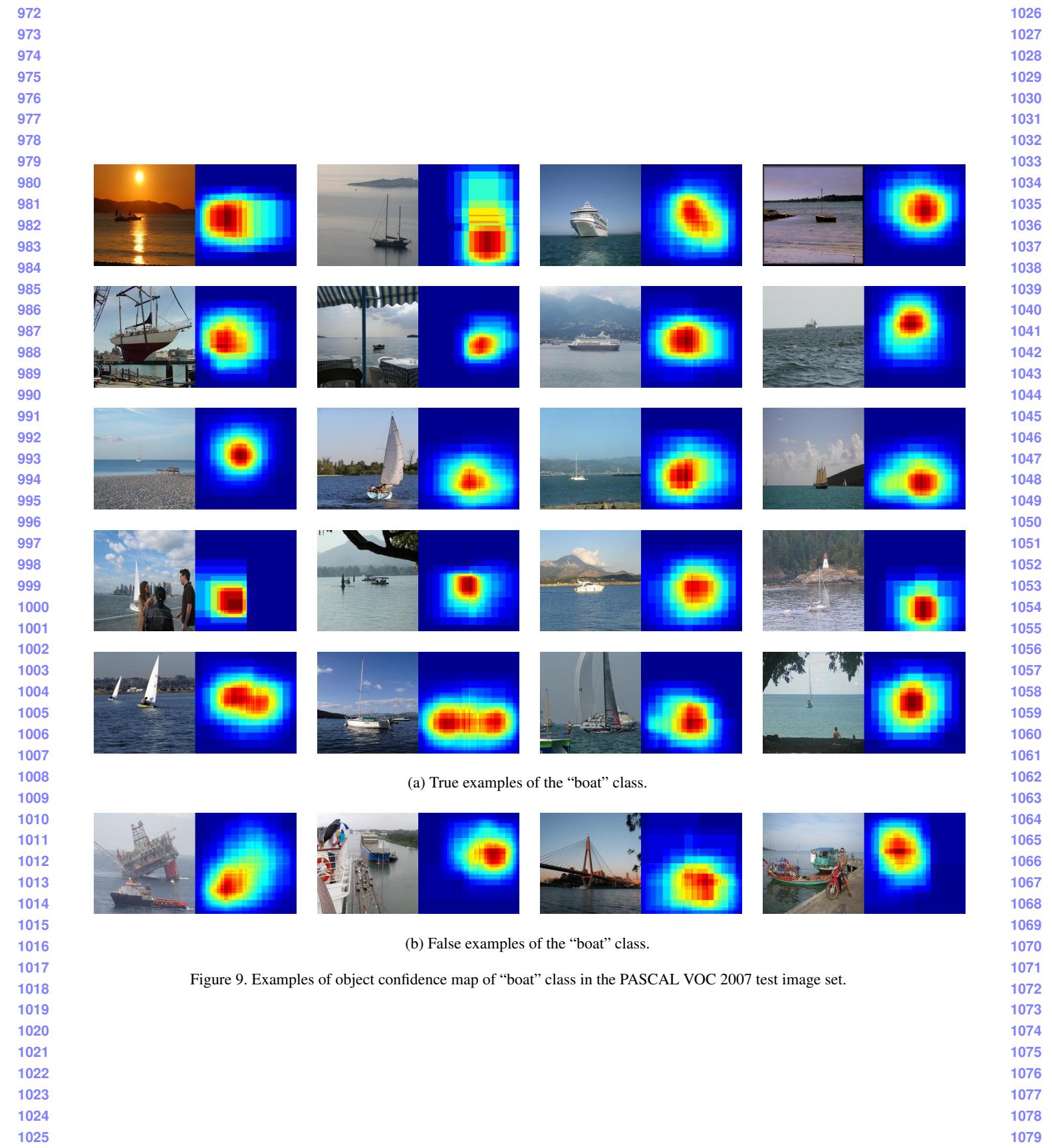
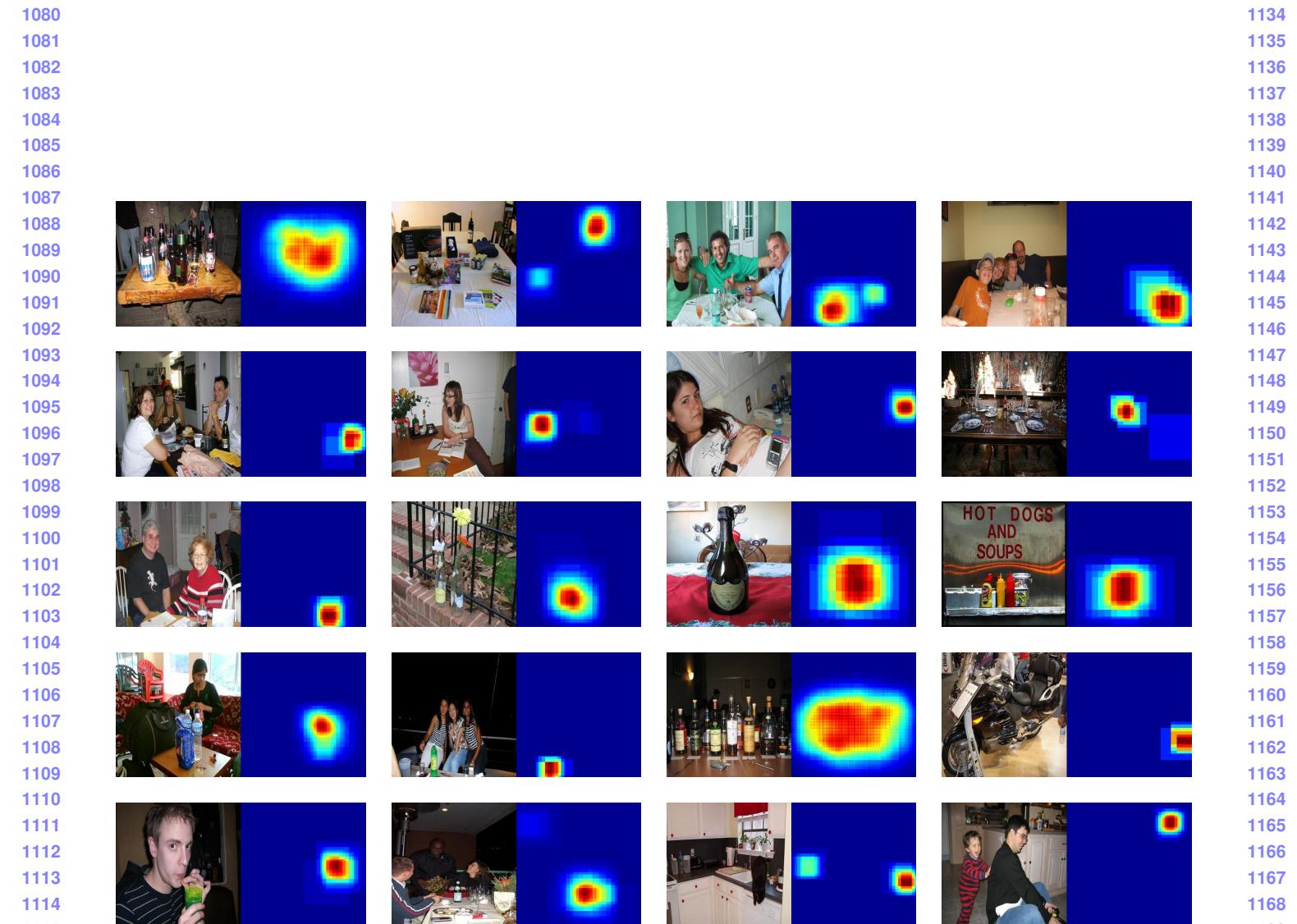
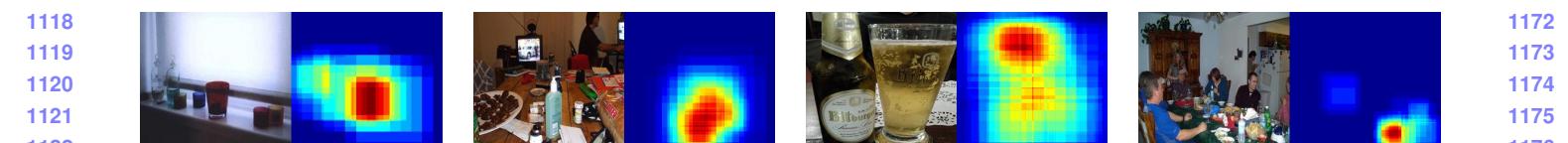


Figure 9. Examples of object confidence map of "boat" class in the PASCAL VOC 2007 test image set.



(a) True examples of the “bottle” class.



(b) False examples of the “bottle” class.

Figure 10. Examples of object confidence map of “bottle” class in the PASCAL VOC 2007 test image set.

1125
1126
1127
1128
1129
1130
1131
1132
11331179
1180
1181
1182
1183
1184
1185
1186
1187

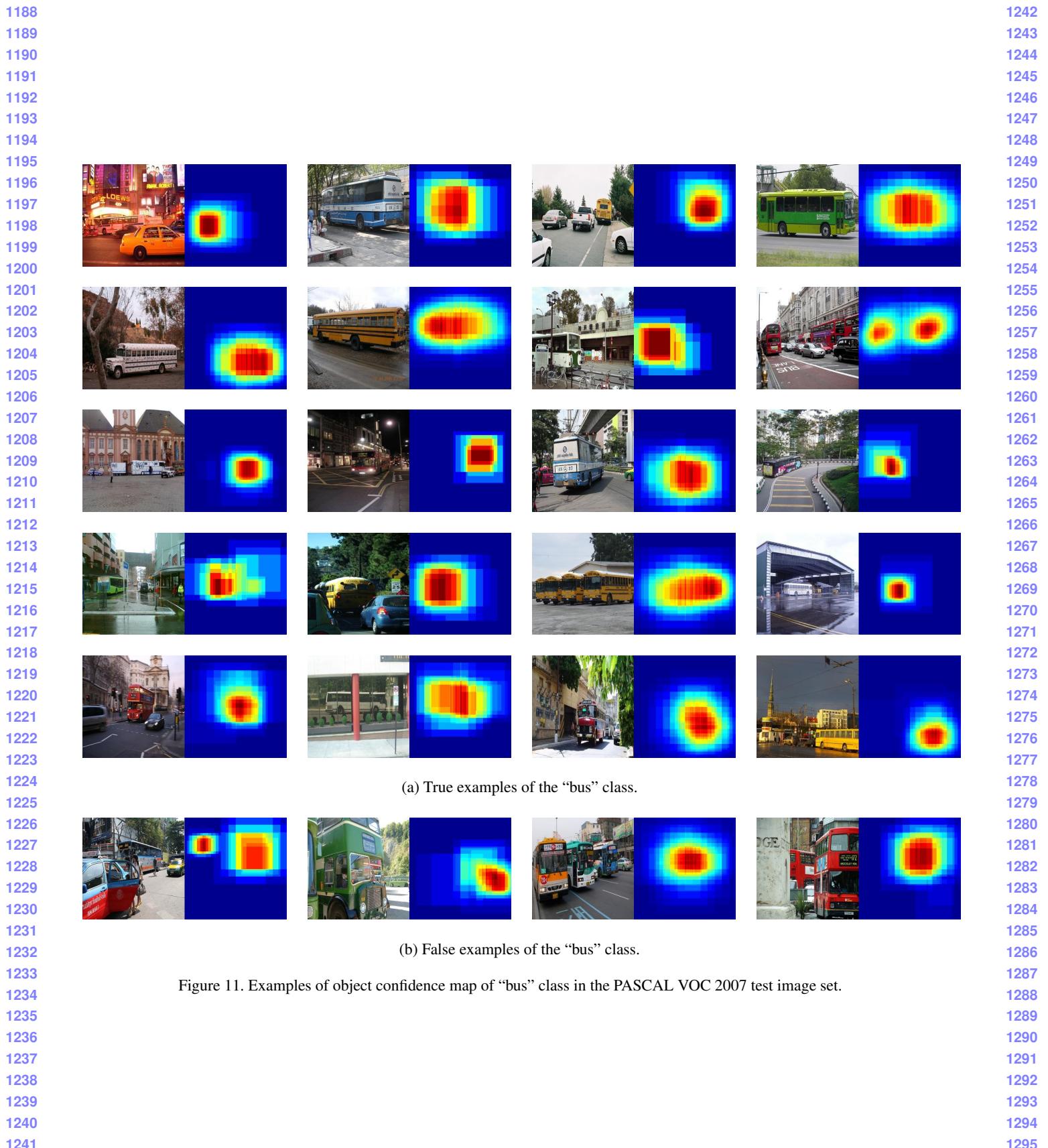


Figure 11. Examples of object confidence map of "bus" class in the PASCAL VOC 2007 test image set.

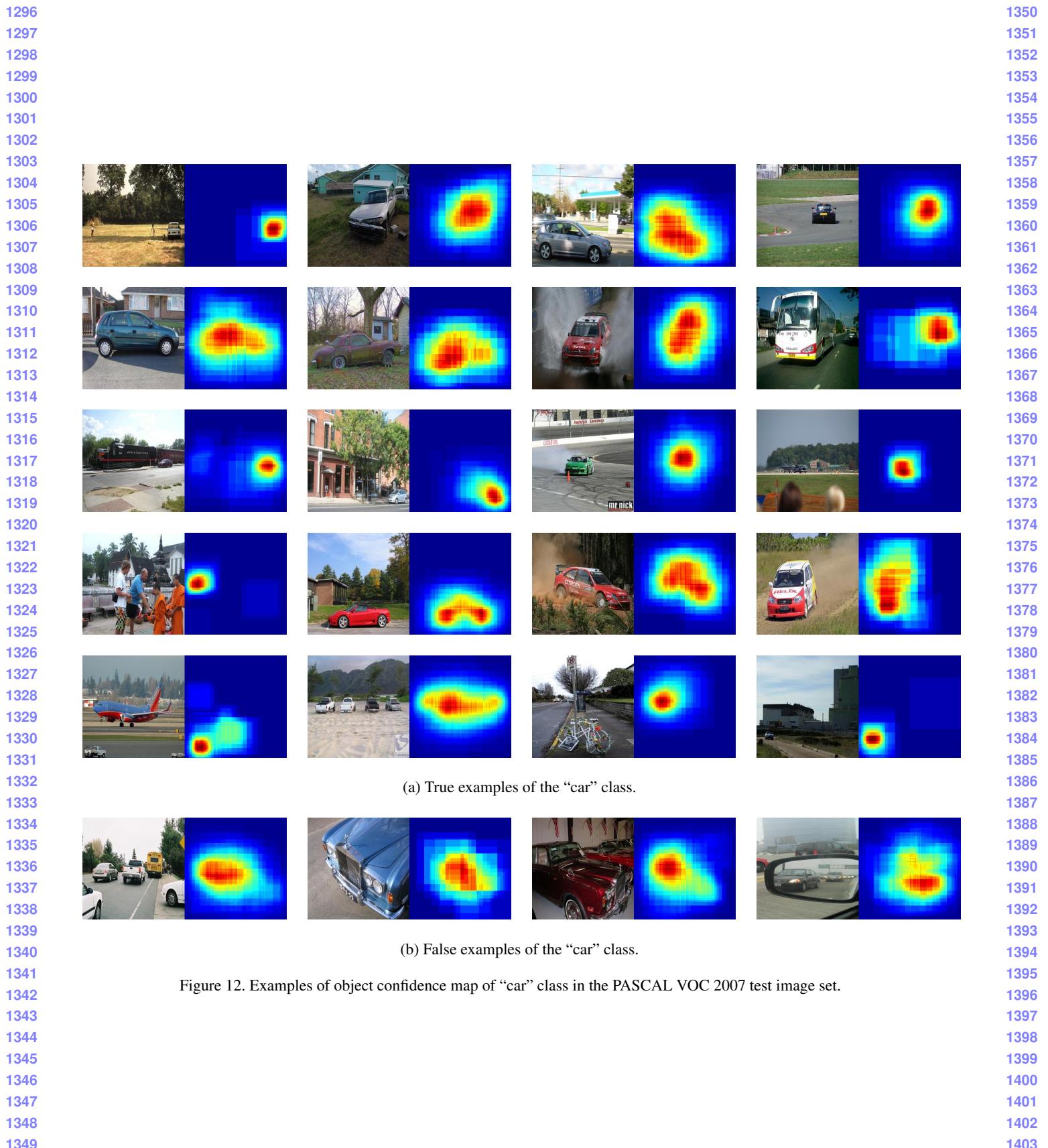


Figure 12. Examples of object confidence map of "car" class in the PASCAL VOC 2007 test image set.

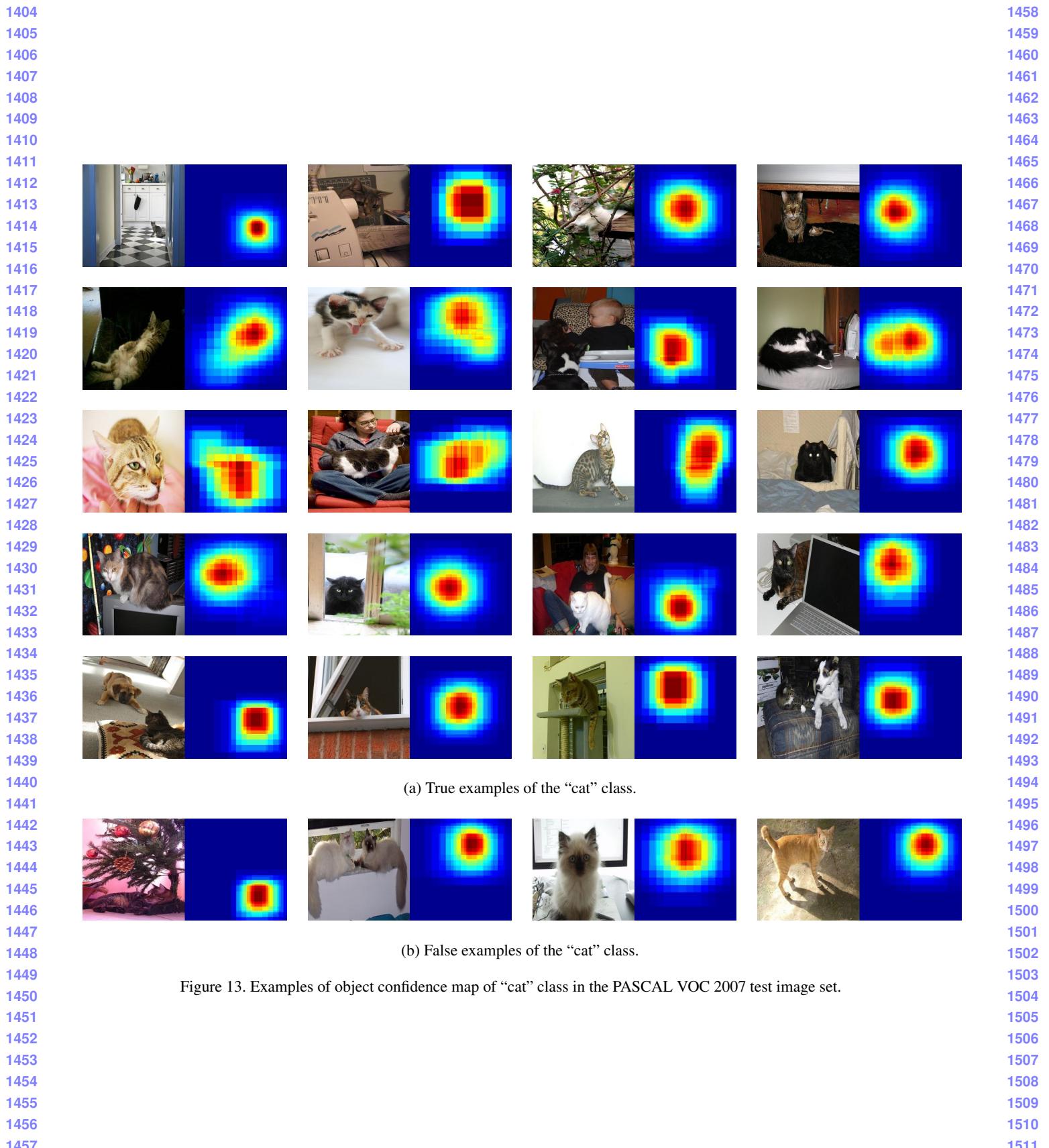
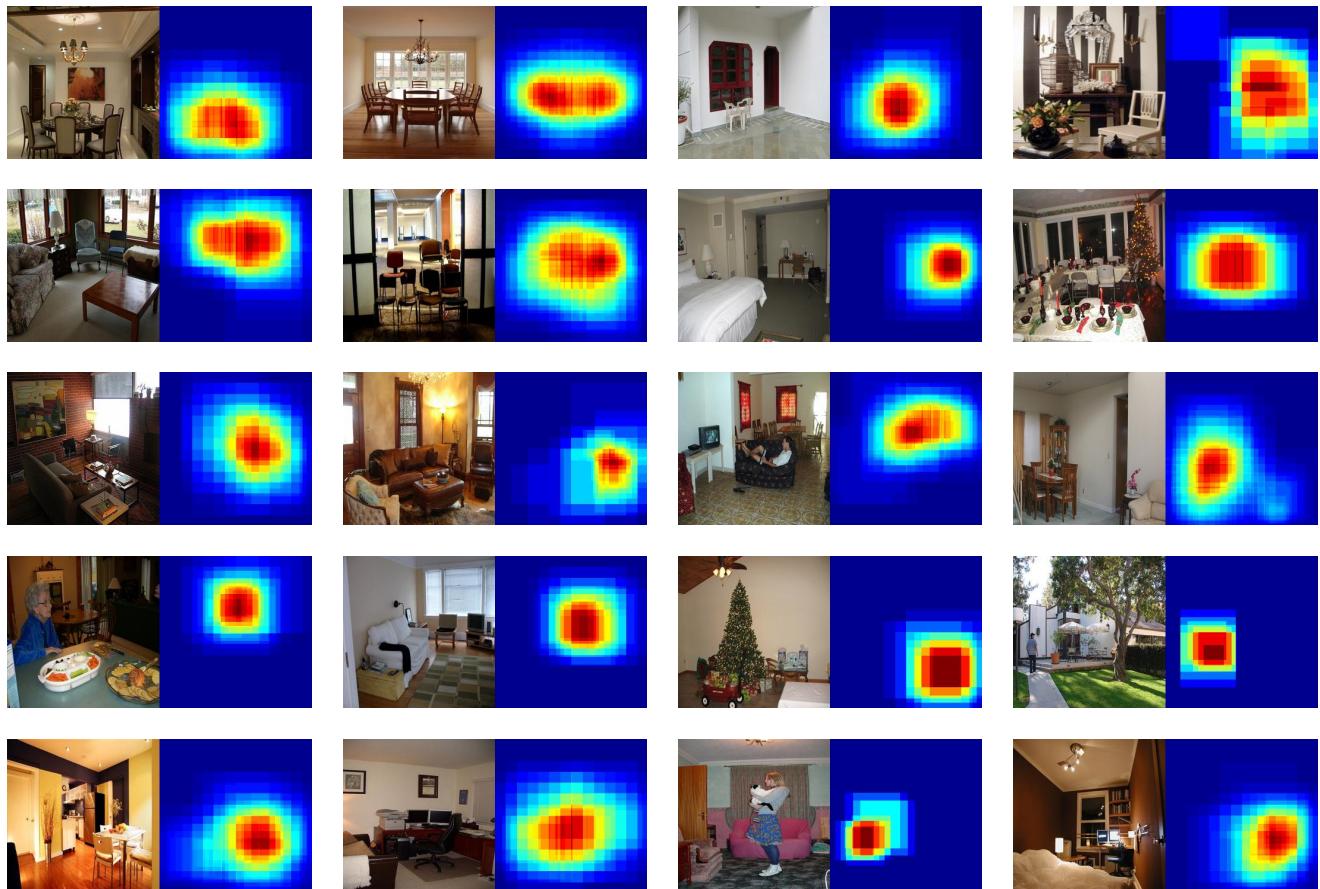


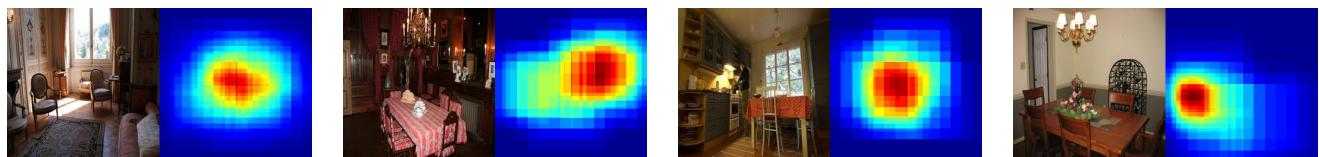
Figure 13. Examples of object confidence map of "cat" class in the PASCAL VOC 2007 test image set.

1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524



(a) True examples of the “chair” class.

1549

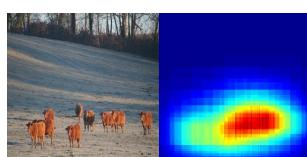
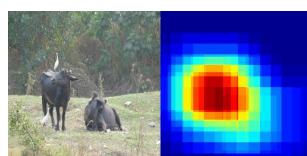
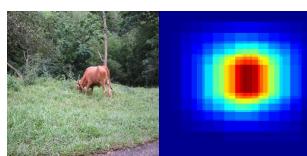
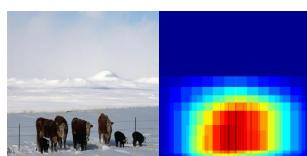
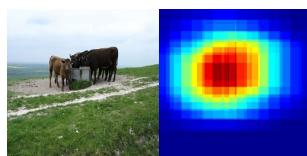
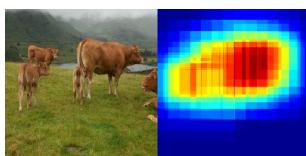
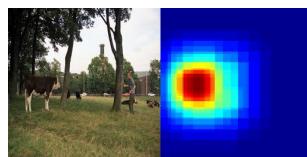
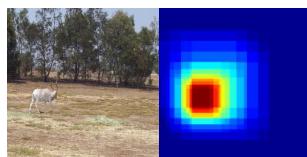
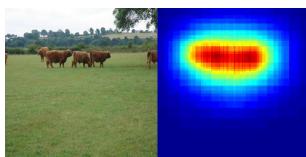
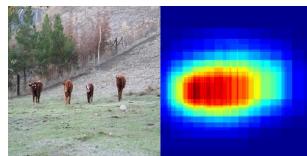
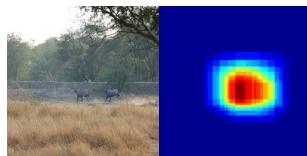
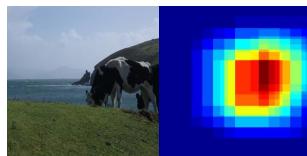
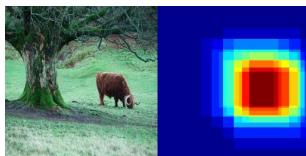


(b) False examples of the “chair” class.

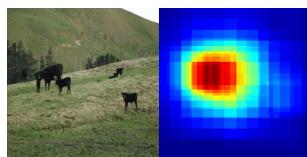
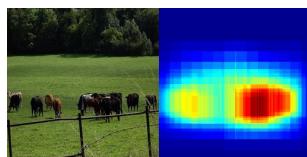
1557
1558
1559
1560
1561
1562
1563
1564
1565

Figure 14. Examples of object confidence map of “chair” class in the PASCAL VOC 2007 test image set.

1620
1621
1622
1623
1624
1625
1626
1627



(a) True examples of the “cow” class.

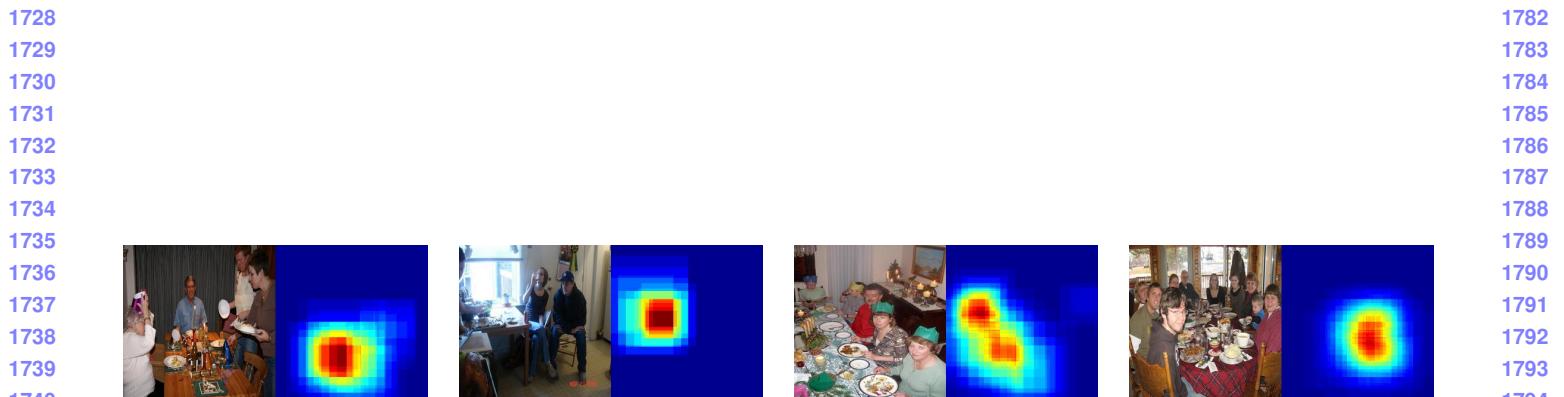


(b) False examples of the “cow” class.

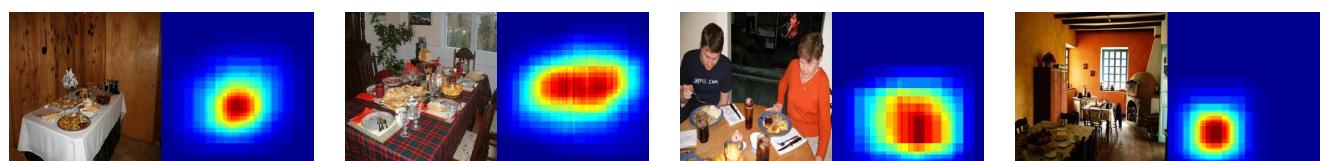
Figure 15. Examples of object confidence map of “cow” class in the PASCAL VOC 2007 test image set.

1665
1666
1667
1668
1669
1670
1671
1672
1673

1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727



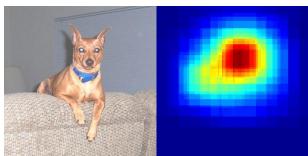
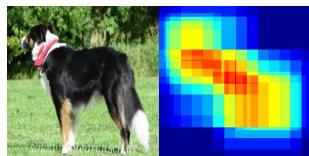
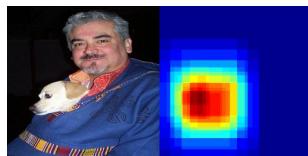
(a) True examples of the “diningtable” class.



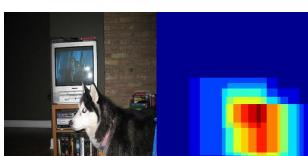
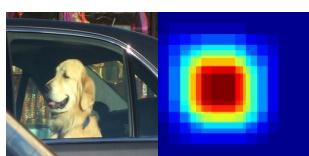
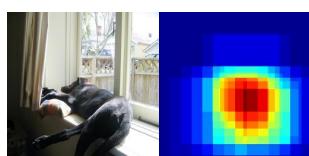
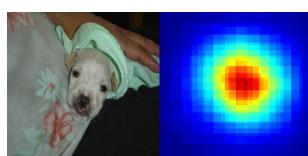
(b) False examples of the “diningtable” class.

Figure 16. Examples of object confidence map of “diningtable” class in the PASCAL VOC 2007 test image set.

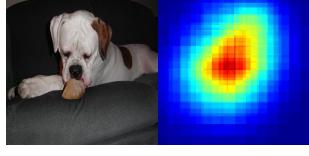
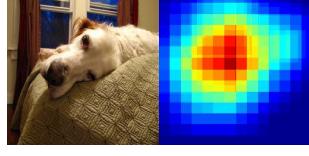
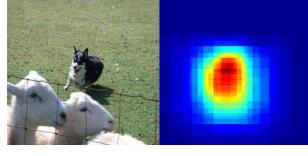
1836
1837
1838
1839
1840
1841
1842
1843



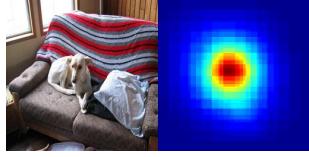
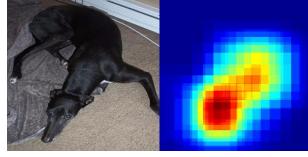
1844
1845
1846
1847
1848
1849



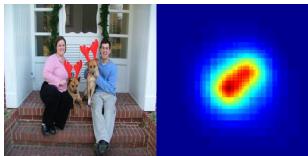
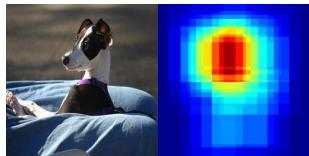
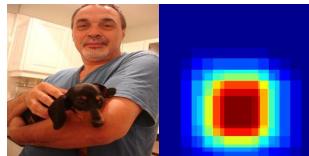
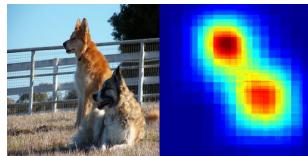
1850



1855
1856
1857
1858
1859
1860



1861
1862
1863
1864
1865

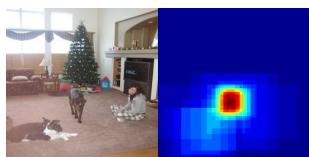
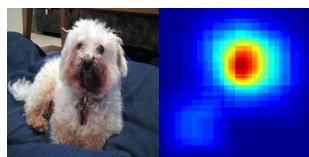
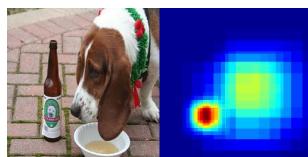


1866

1871

(a) True examples of the “dog” class.

1872



1873

(b) False examples of the “dog” class.

1874

1881

Figure 17. Examples of object confidence map of “dog” class in the PASCAL VOC 2007 test image set.

1882

1883

1884

1885

1886

1887

1888

1889

1890

1891

1892

1893

1894

1895

1896

1897

1898

1899

1900

1901

1902

1903

1904

1905

1906

1907

1908

1909

1910

1911

1912

1913

1914

1915

1916

1917

1918

1919

1920

1921

1922

1923

1924

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

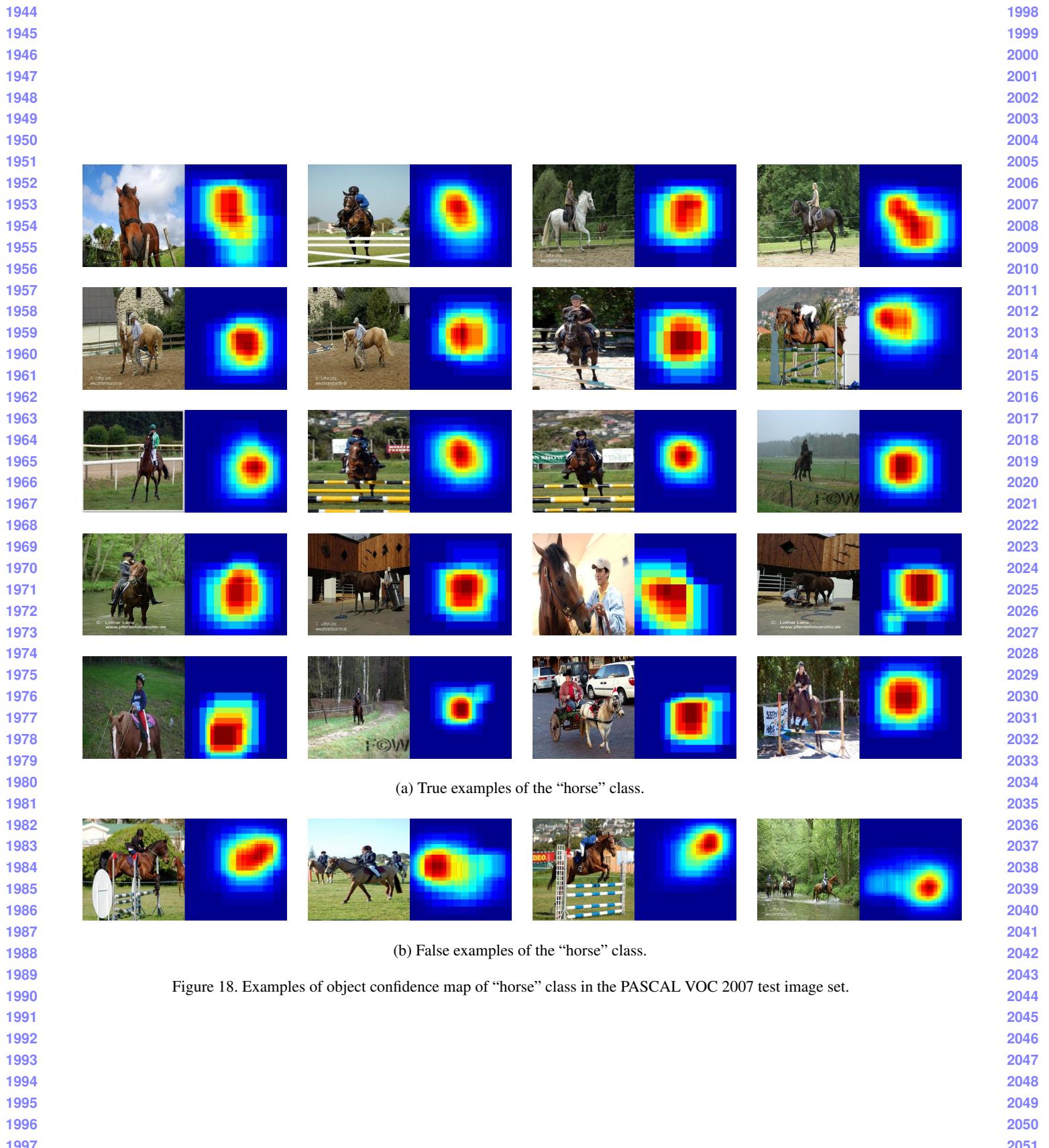


Figure 18. Examples of object confidence map of "horse" class in the PASCAL VOC 2007 test image set.

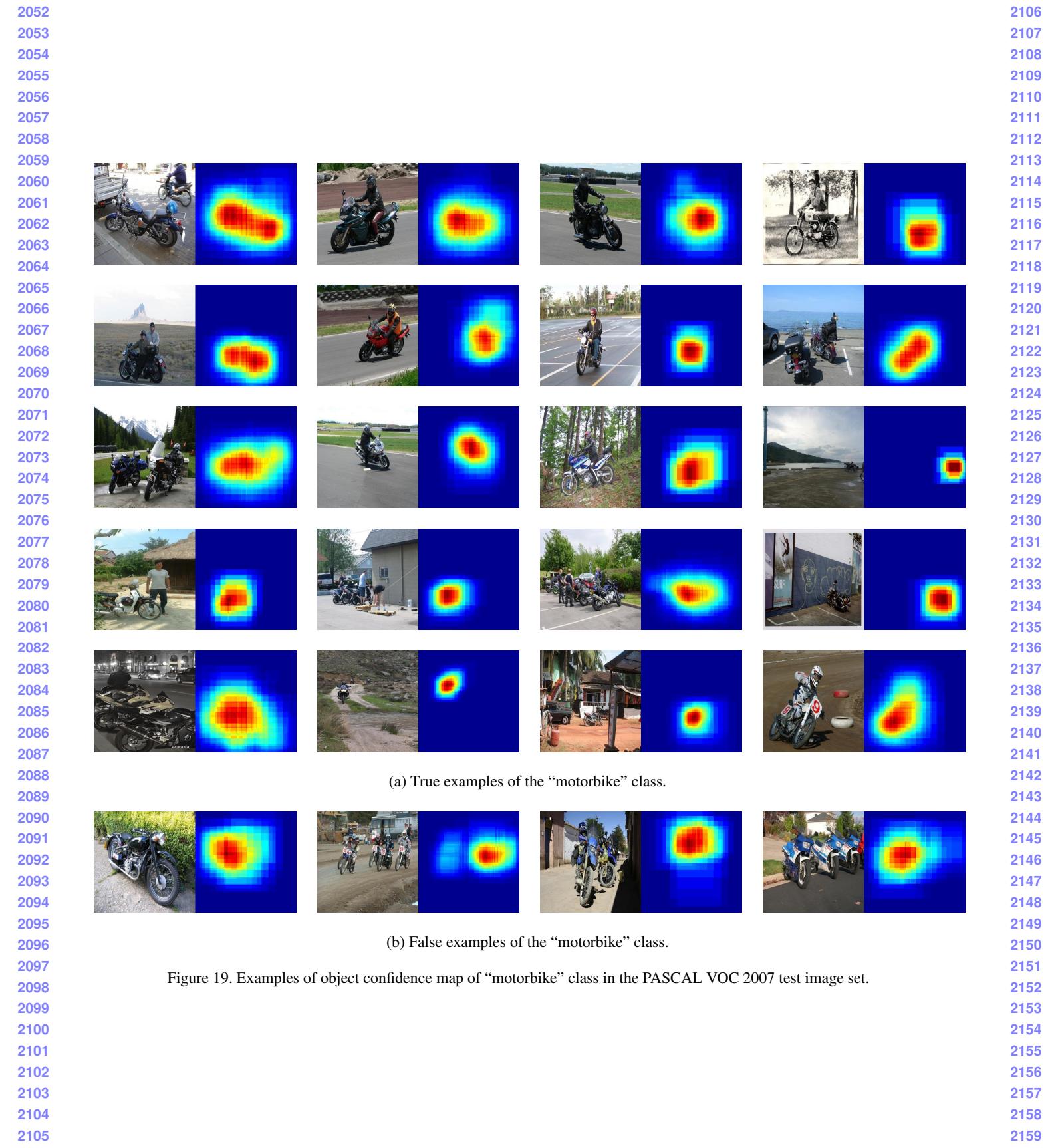
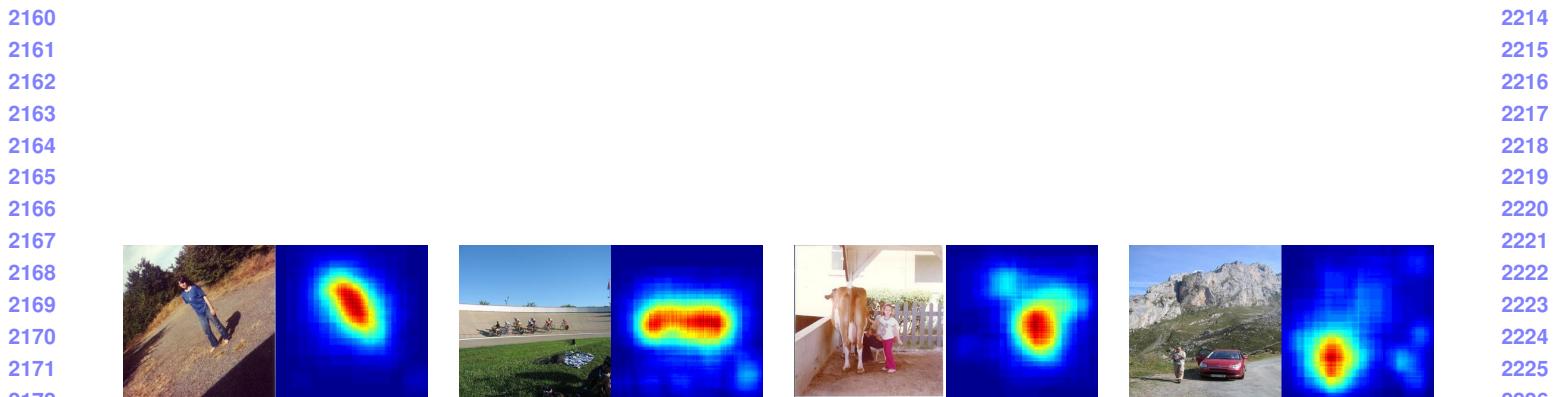
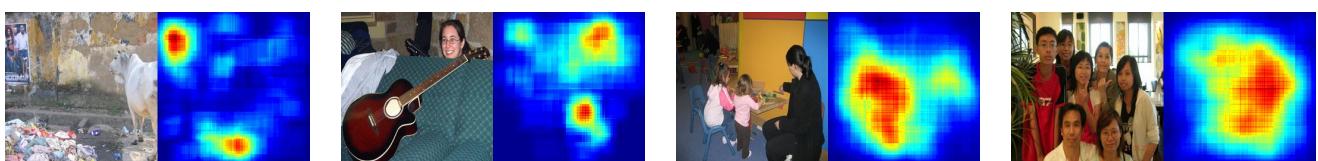


Figure 19. Examples of object confidence map of "motorbike" class in the PASCAL VOC 2007 test image set.



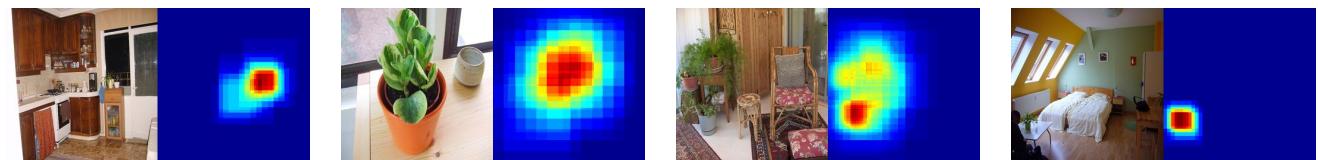
(a) True examples of the “person” class.



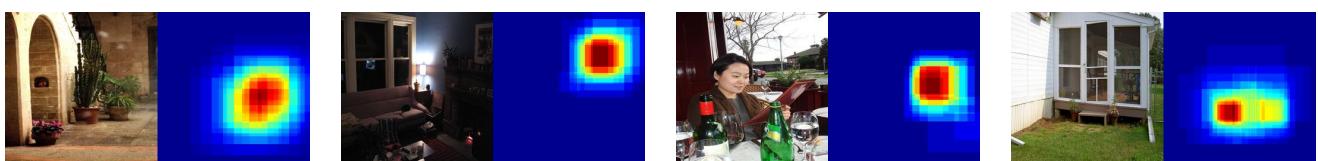
(b) False examples of the “person” class.

Figure 20. Examples of object confidence map of “person” class in the PASCAL VOC 2007 test image set.

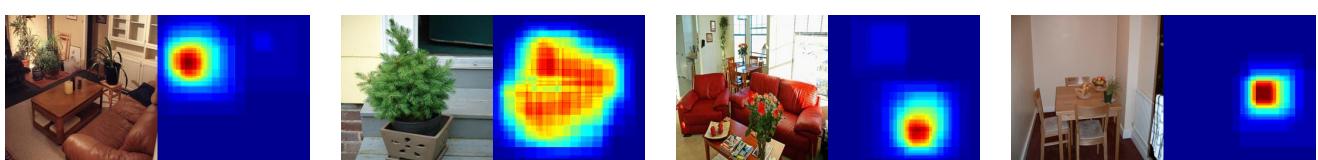
2268			2322
2269			2323
2270			2324
2271			2325
2272			2326
2273			2327
2274			2328
2275			2329
2276			2330
2277			2331
2278			2332
2279			2333
2280			2334



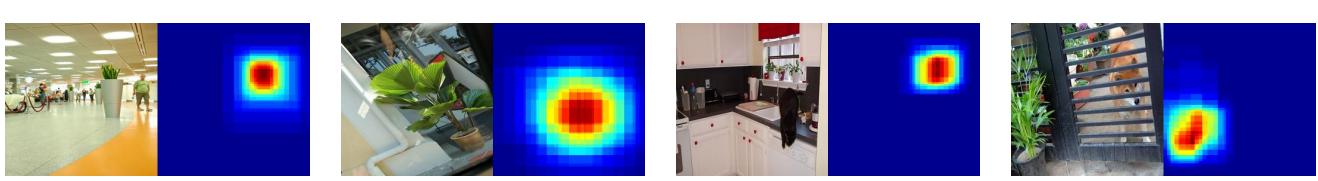
2281			2335
2282			2336
2283			2337
2284			2338
2285			2339



The figure displays five pairs of images arranged horizontally. Each pair consists of a photograph on the left and a heatmap on the right. The first pair shows a living room with a sofa and a small table; the heatmap highlights a central area around the sofa. The second pair shows a potted plant; the heatmap shows a multi-peaked pattern centered on the plant. The third pair shows a living room with red leather couches; the heatmap highlights a central area around the couches. The fourth pair shows a dining room with a table and chairs; the heatmap highlights a central area around the table. The fifth pair shows a hallway; the heatmap highlights a central area near the entrance.

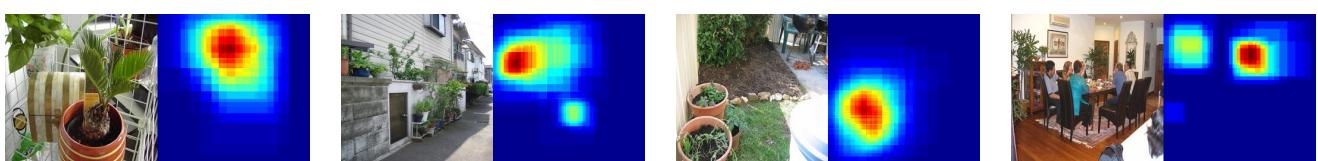


This row contains five pairs of images. Each pair consists of a photograph on the left and a heatmap on the right. The first pair shows a hallway with people walking; the heatmap highlights a central area. The second pair shows a potted plant; the heatmap highlights the center of the plant. The third pair shows a kitchen; the heatmap highlights the center of the room. The fourth pair shows a window with a grid pattern; the heatmap highlights the center of the window frame. The fifth pair shows a potted plant again; the heatmap highlights the center of the plant.



(a) True examples of the “pottedplant” class.

This row contains five pairs of images. From left to right: 1) A potted plant on a shelf with a heatmap overlay showing a central peak of high intensity. 2) An outdoor scene of a garden with a heatmap overlay showing a central peak. 3) A potted plant on a lawn with a heatmap overlay showing a central peak. 4) An indoor dining room scene with several people at a table, with a heatmap overlay showing a central peak. 5) An outdoor scene of a garden with a heatmap overlay showing a central peak.



(b) False examples of the “pottedplant” class.

Figure 21. Examples of object confidence map of “pottedplant” class in the PASCAL VOC 2007 test image set.

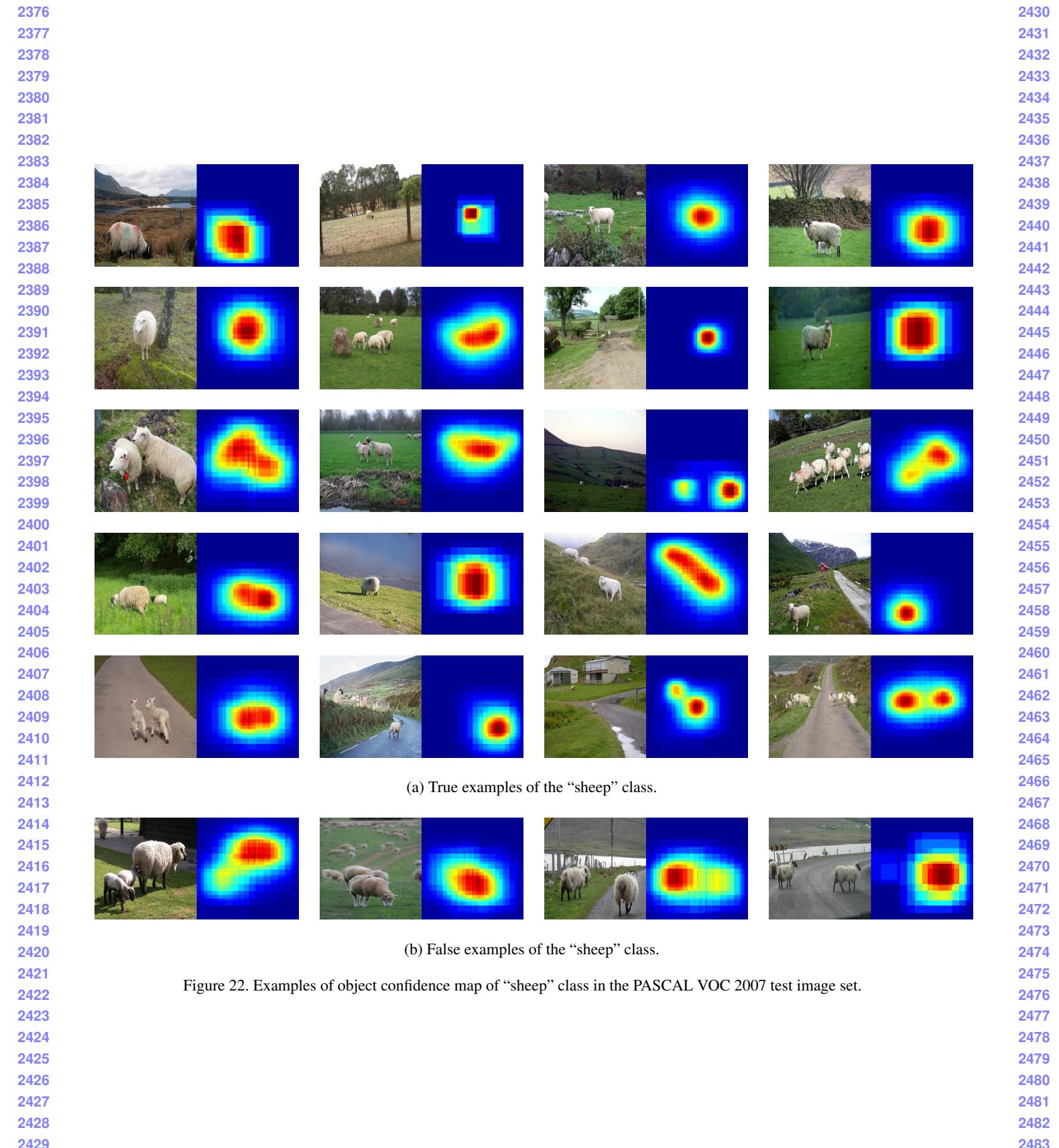


Figure 22. Examples of object confidence map of “sheep” class in the PASCAL VOC 2007 test image set.

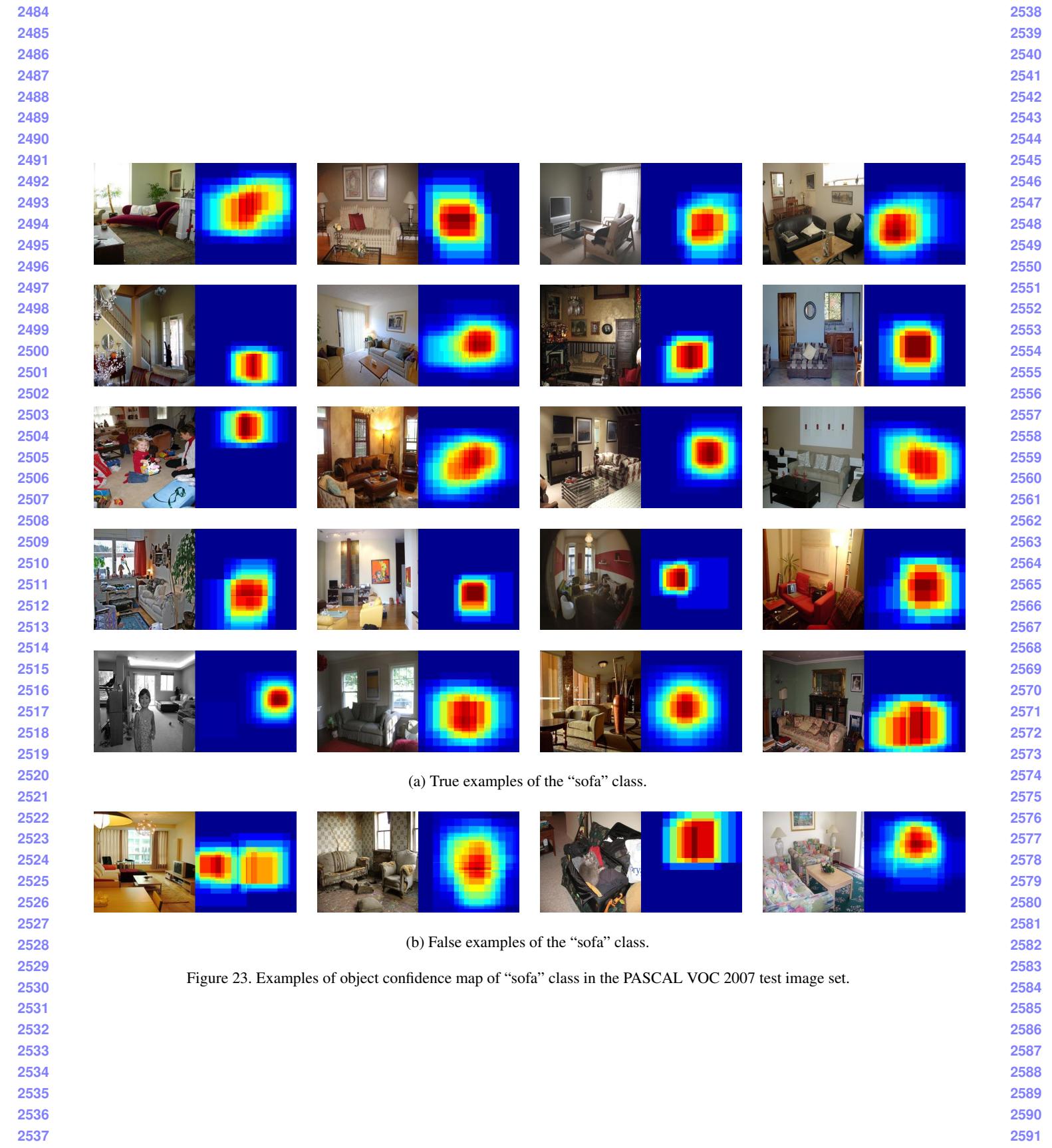
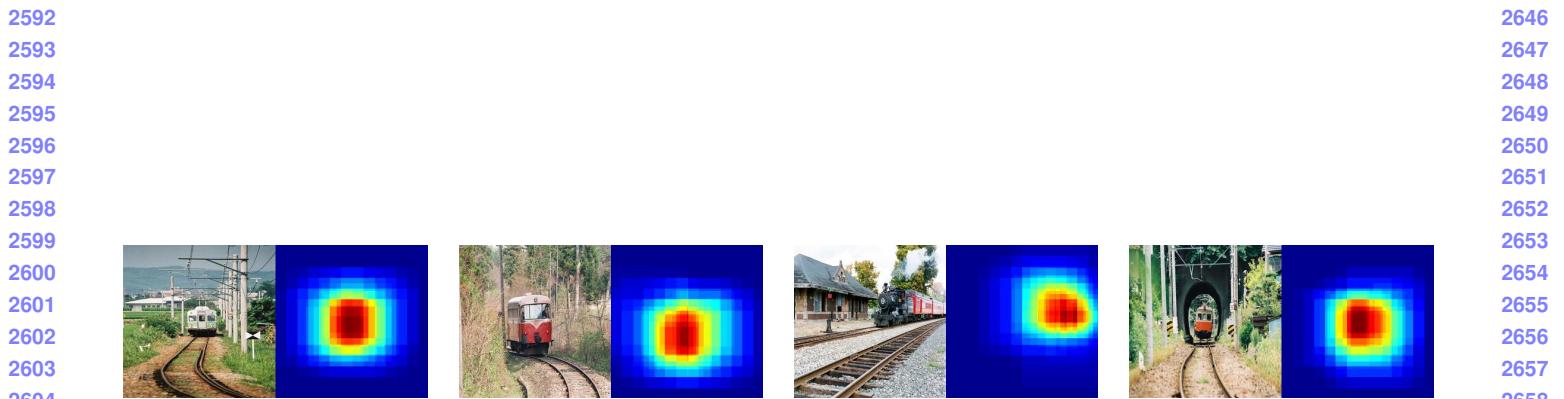
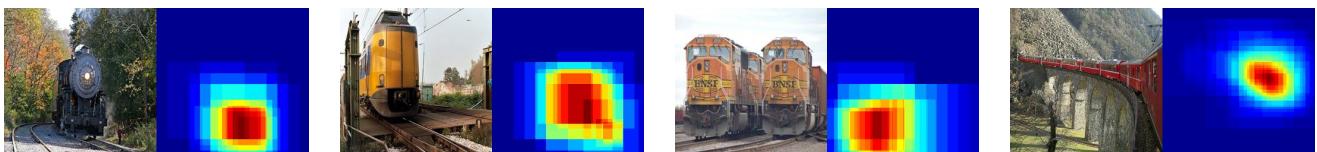


Figure 23. Examples of object confidence map of “sofa” class in the PASCAL VOC 2007 test image set.



(a) True examples of the “train” class.



(b) False examples of the “train” class.

Figure 24. Examples of object confidence map of “train” class in the PASCAL VOC 2007 test image set.

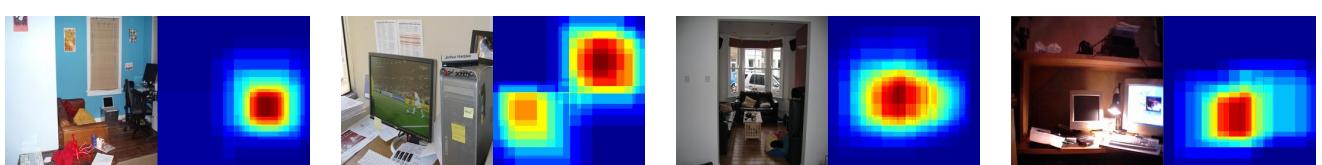
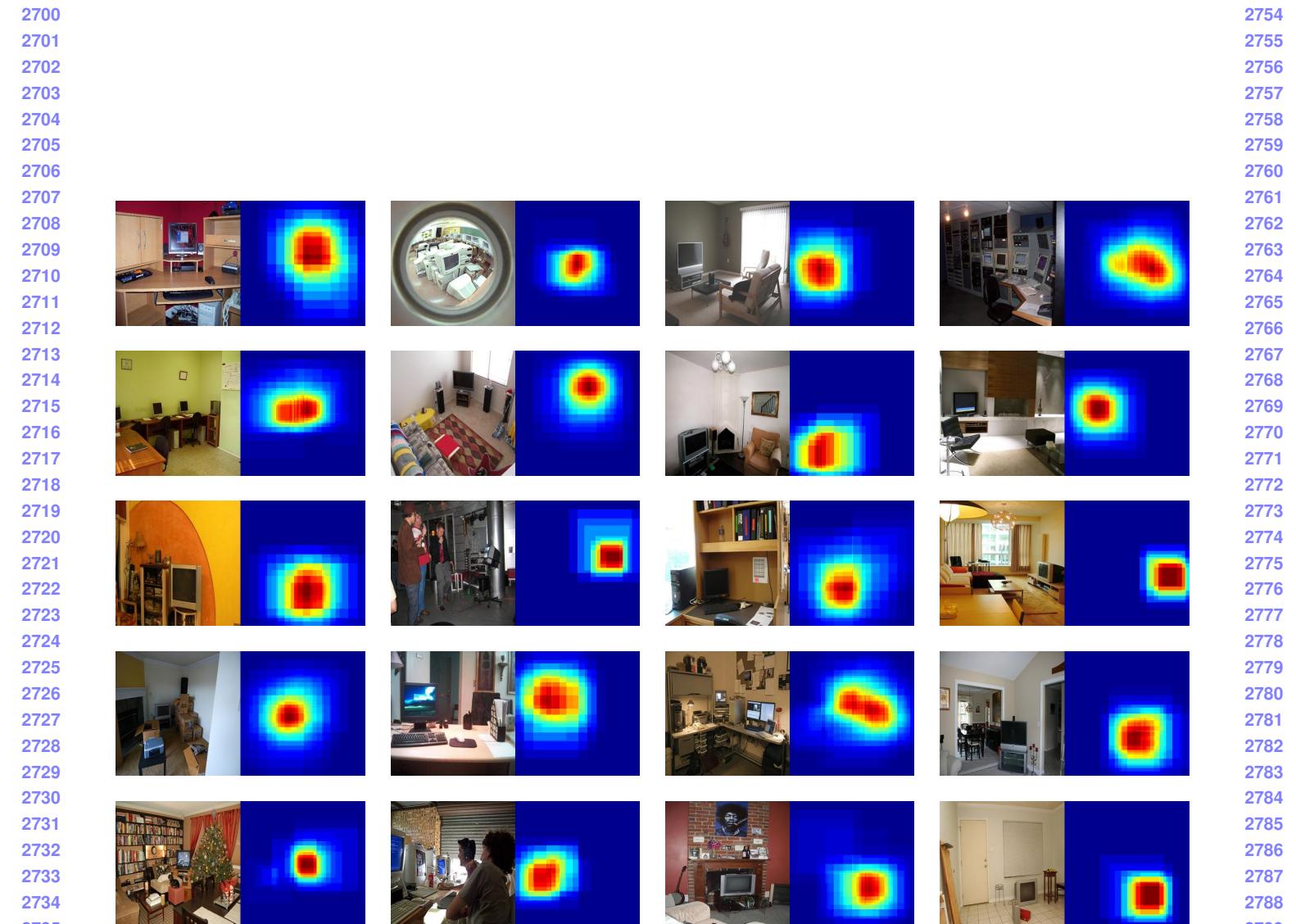


Figure 25. Examples of object confidence map of “tvmonitor” class in the PASCAL VOC 2007 test image set.

2808	References	2862
2809		2863
2810	[1] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture	2864
2811	for fast feature embedding. <i>arXiv preprint arXiv:1408.5093</i> , 2014. 1, 2, 3	2865
2812	[2] A. Vedaldi and K. Lenc. MatConvNet: Convolutional neural networks for matlab. http://www.vlfeat.org/matconvnet/ ,	2866
2813	2014. 1	2867
2814		2868
2815		2869
2816		2870
2817		2871
2818		2872
2819		2873
2820		2874
2821		2875
2822		2876
2823		2877
2824		2878
2825		2879
2826		2880
2827		2881
2828		2882
2829		2883
2830		2884
2831		2885
2832		2886
2833		2887
2834		2888
2835		2889
2836		2890
2837		2891
2838		2892
2839		2893
2840		2894
2841		2895
2842		2896
2843		2897
2844		2898
2845		2899
2846		2900
2847		2901
2848		2902
2849		2903
2850		2904
2851		2905
2852		2906
2853		2907
2854		2908
2855		2909
2856		2910
2857		2911
2858		2912
2859		2913
2860		2914
2861		2915