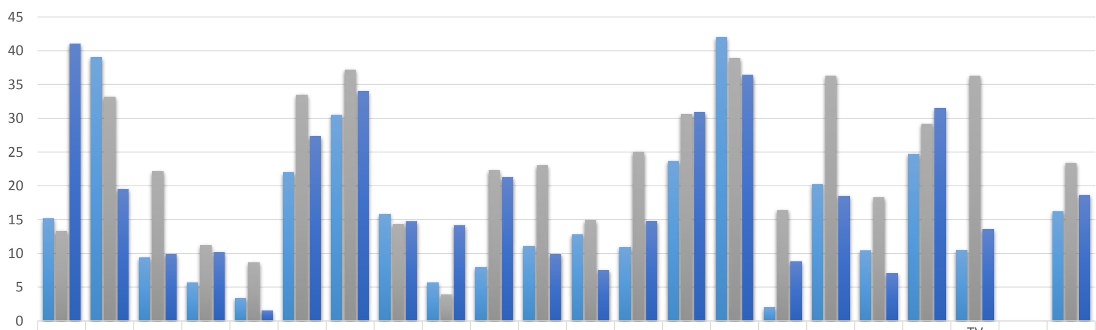
Supplementary material: Backtracking ScSPM Image Classifier for Weakly Supervised Top-down Saliency

Hisham Cholakkal Jubin Johnson Deepu Rajan
Nanyang Technological University Singapore
{hisham002, jubin001, asdrajan}@ntu.edu.sg

1. Quantitaive Comparison of proposed weakly supervised approach vs fully supervised top-down saliency approaches-PASCAL VOC-07 (patch-level precision rates at EER (%))



	Aeroplane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow	Dining table	Dog	Horse	Motor bike	Person	Potted plant	Sheep	Sofa	Train	TV Monitor		Mean
Yang and Yang [36]	15.2	39	9.4	5.7	3.4	22	30.5	15.8	5.7	8	11.1	12.8	10.9	23.7	42	2	20.2	10.4	24.7	10.5		16.2
LCCSC [5]	13.3	33.2	22.1	11.2	8.6	33.5	37.2	14.3	3.9	22.3	23	14.9	25	30.6	38.9	16.4	36.3	18.3	29.2	36.3		23.4
Proposed	41	19.5	9.9	10.2	1.5	27.3	34	14.7	14.1	21.2	9.9	7.5	14.8	30.9	36.4	8.8	18.5	7.1	31.5	13.6		18.6

Figure 1. Patch-level precision rates at EER (%) on PASCAL VOC-07 dataset.

2. Quantitative comparison of pixel labeling accuracy with fully supervised object class-segmentation methods in PASCAL VOC-07

Table 1. Pixel classification accuracy of individual classes as% of correctly labeled pixels.

Background	Aero plane	Bicycle	Bird	Boat	Bottle	Bus	Car	Cat	Chair	Cow
73.05	17.9	35.3	5.8	9.1	0.25	15.8	40.0	7.4	18.9	3.12

Dining table	Dog	Horse	Motor bike	Person	Potted Plant	Sheep	Sofa	Train	TV Monitor	Average Accuracy	Overall % of pixels correctly labeled
1.3	6.9	15.13	6.1	37.2	13.8	7.0	9.3	20.7	16.3	17.07	57.31

For comparison with fully supervised class segmentation approaches [2, 11], saliency maps corresponding to each object model are applied over a test image. The pixels that have overlapping saliency values of different object classes are assigned the class with the highest saliency. Object-specific saliency maps are thresholded at their EER as explained in the paper. The pixels that fall below this threshold are classified as background. Following such an approach, 57.3% of the total pixels are correctly classified, which is better than 57% in [11], which uses fully supervised learning of CRF on superpixel based image features. The percentage of correctly labeled pixels for each class is evaluated and averaged across 21 classes(20 object classes and background class). We obtain a mean of 17.07%, which is comparable to 16% of [2] using similar parameters (SIFT like IHOG feature, hierarchical K-means based dictionary, using spatial bins and linear classifier).

3. Qualitative comparison with weakly supervised top-down saliency [25]

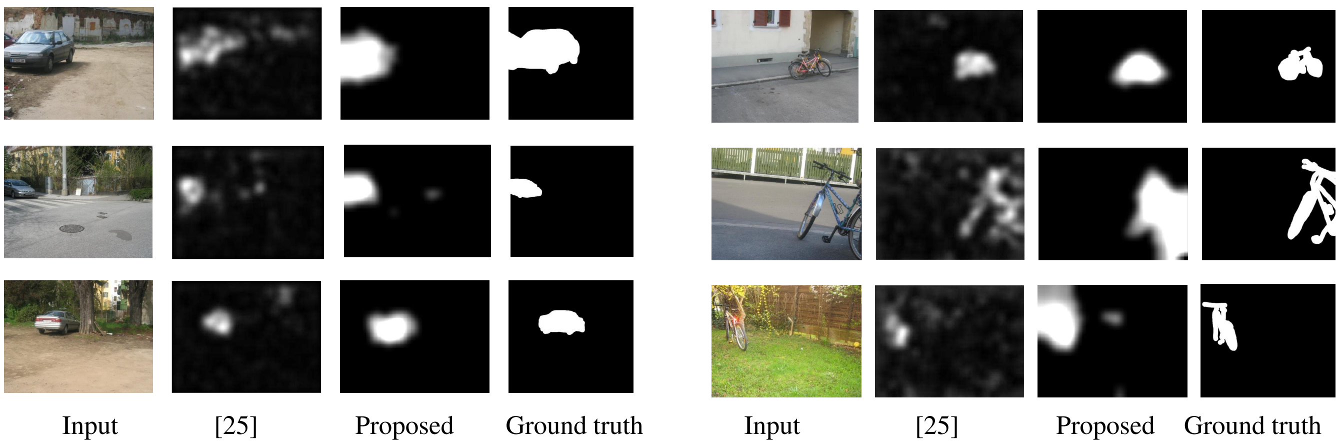


Figure 2. Qualitative comparison with [25].

Being an image categorization approach, [25] does not report quantitative results of saliency estimation. Instead saliency maps corresponding to 6 images from Graz-02 dataset are given in the paper. In contrast to traditional training-test set split, they used 150 even numbered images per category for training and 150 odd numbered images per category for testing. We also evaluated our model using this train-test split and saliency maps corresponding to these 6 test images are produced using our approach. Fig. 2 shows that the proposed method clearly outperforms [25] (they evaluate only on Graz-02 dataset and not on PASCAL VOC-07).

4. Qualitative comparison with fully supervised top-down saliency approaches

Here we compare our saliency maps with [36, 20, 5], in both Graz-02 and PASCAL VOC-07 datasets. Saliency maps of [20, 5] are provided by the authors. Saliency maps of [36] are produced using the code provided by the authors.

4.1. Graz-02

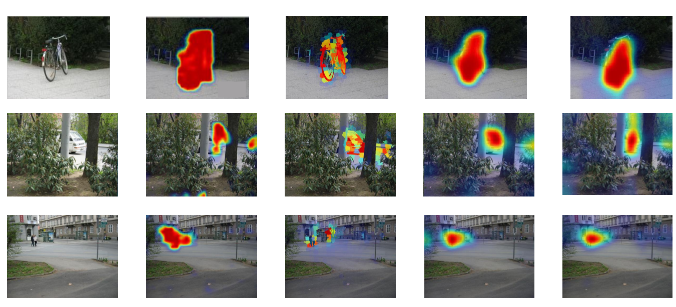


Figure 3. Qualitative comparison of our weakly supervised approach with fully supervised top-down saliency approaches for some images of Graz-02 dataset.

4.2. PASCAL VOC-07

Qualitative comparison with fully supervised top-down saliency approaches is given in Fig. 4.

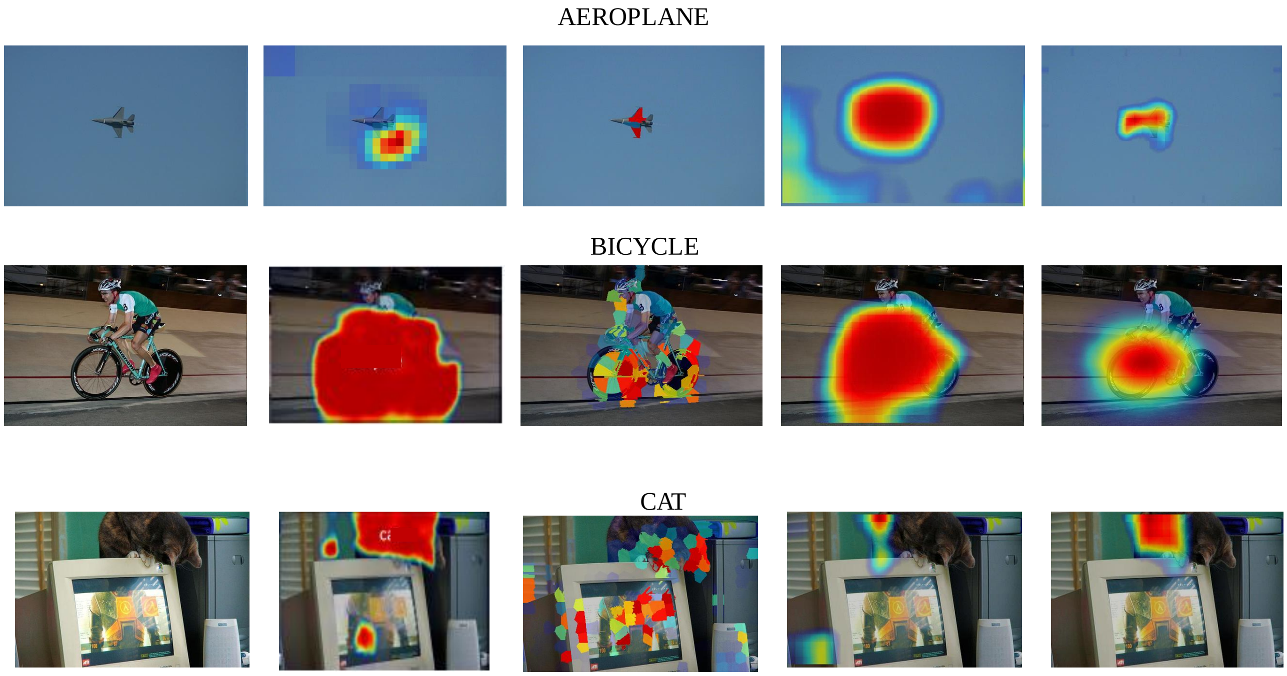


Figure 4. Qualitative comparison of the proposed weakly supervised approach against fully supervised top-down saliency approaches for PASCAL VOC-07.

5. More Saliency maps by the proposed method on Graz-02 and PASCAL VOC-07

5.1. Graz-02

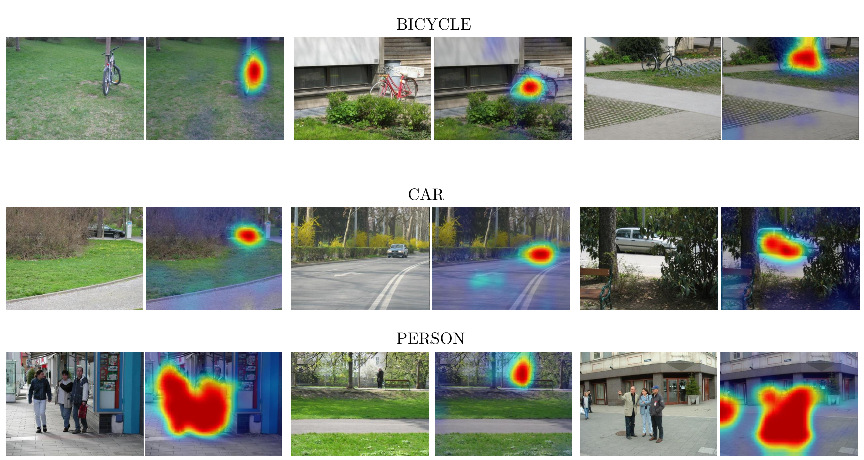


Figure 5. Saliency maps generated using the proposed weakly supervised approach on Graz-02 images.

5.2. PASCAL VOC-07

Saliency maps generated on PASCAL VOC-07 segmentation and detection test images are shown in Fig. 6 , 7.

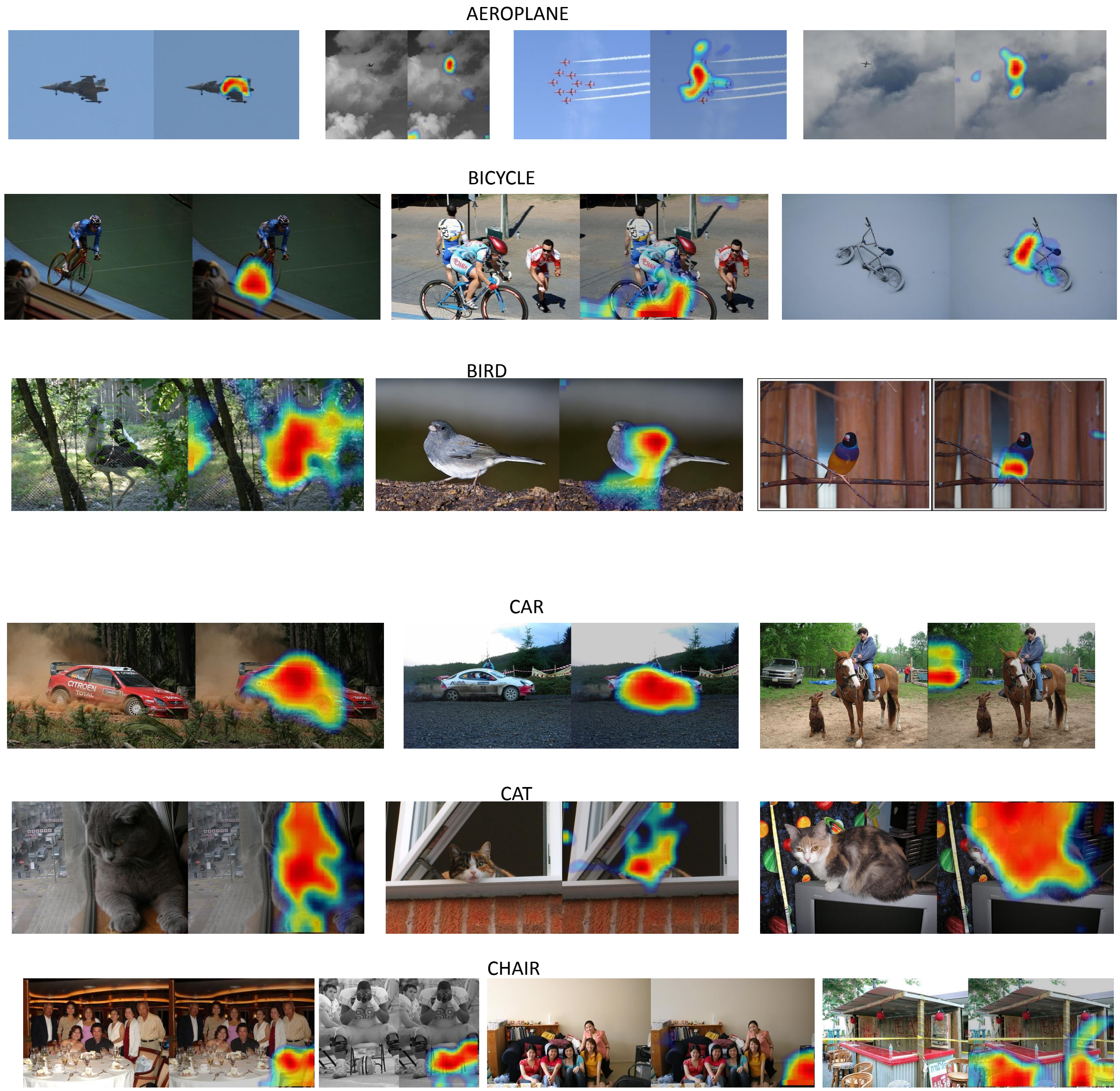


Figure 6. Qualitative results of proposed weakly supervised approach on PASCAL VOC-07. It is hard to identify the presence of bird (column 1) inbetween trees. Car (column 3) is successfully identified even with the presence of other 3 object categories-dog,horse and person.

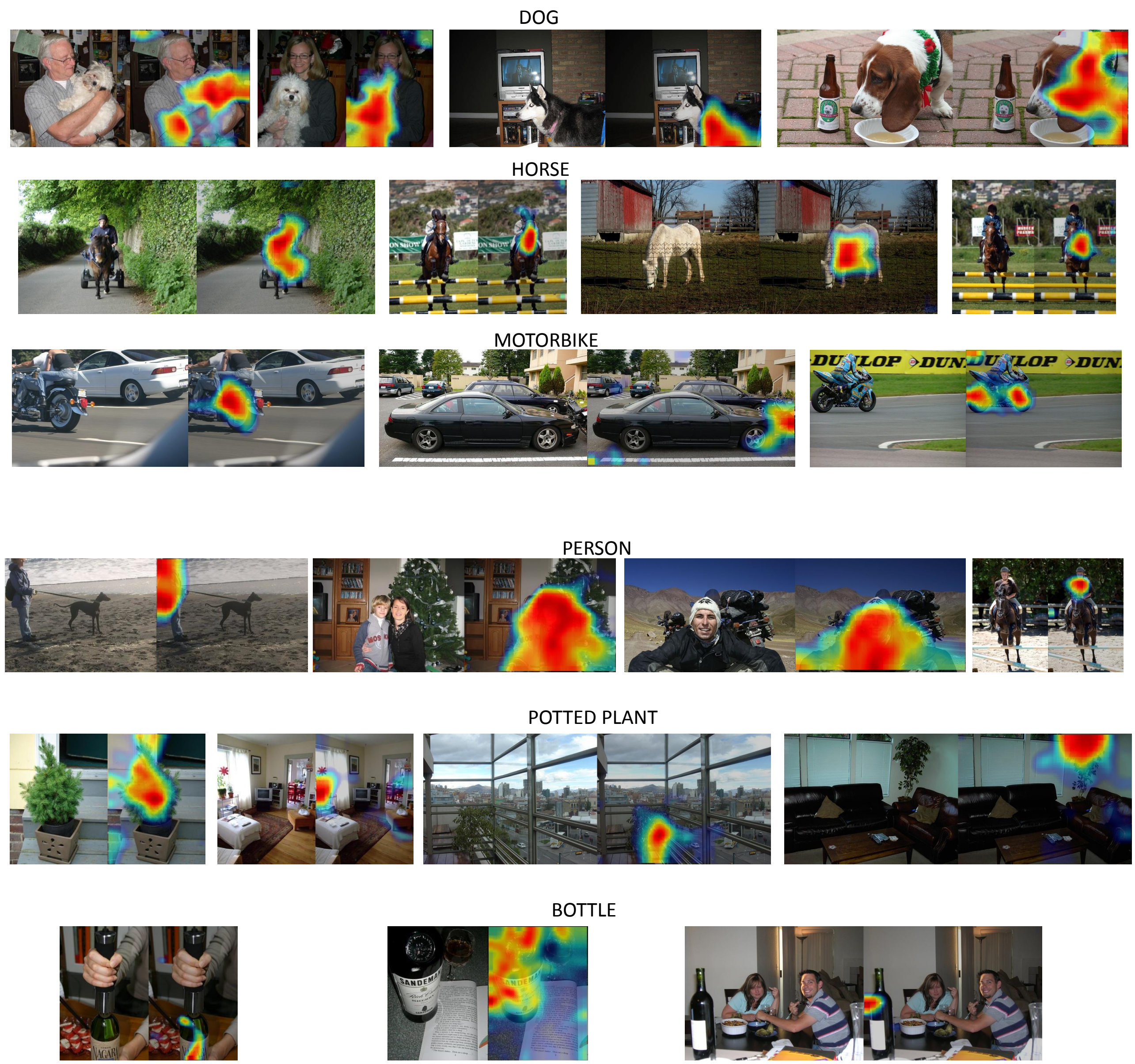


Figure 7. Qualitative results of proposed weakly supervised approach on PASCAL VOC-07. Motorbike is correctly identified, even-though it is occluded with car (column 2). It is hard to identify the presence of potted-plant which is correctly marked as salient (Column 2)

6. Class-specific patches extracted using R-ScSPM saliency

Image patches having R-ScSPM saliency ≥ 0.5 are considered as the class-specific patches of that image. Here we show the class-specific patches extracted on Graz-02 and PASCAL VOC-07 images.

6.1. Class-specfic patches extracted on Graz-02 images

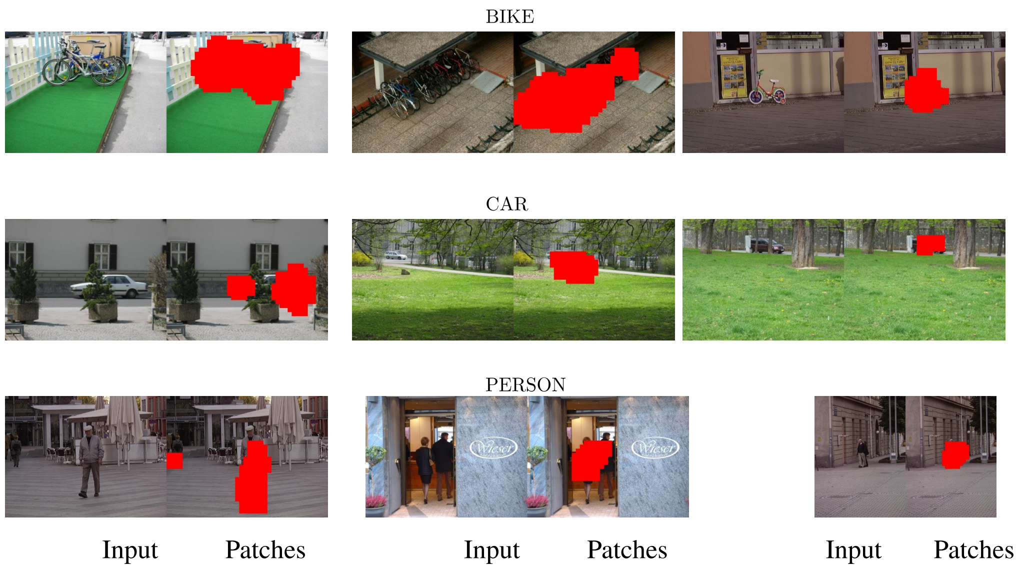


Figure 8. Class-specific patches (red) identified on training images by thresholding R-ScSPM saliency at *0.5*. In the car image (column 1), car patches are selected by removing the tree patches that occlude the car.

6.1.1 Class-specific patches extracted on PASCAL VOC-07 images

Fig. 9, 10 shows the class-specific patches extracted on PASCAL VOC-07 detection images.

Figure 9. Class-specific patches identified on PASCAL VOC-07 images.

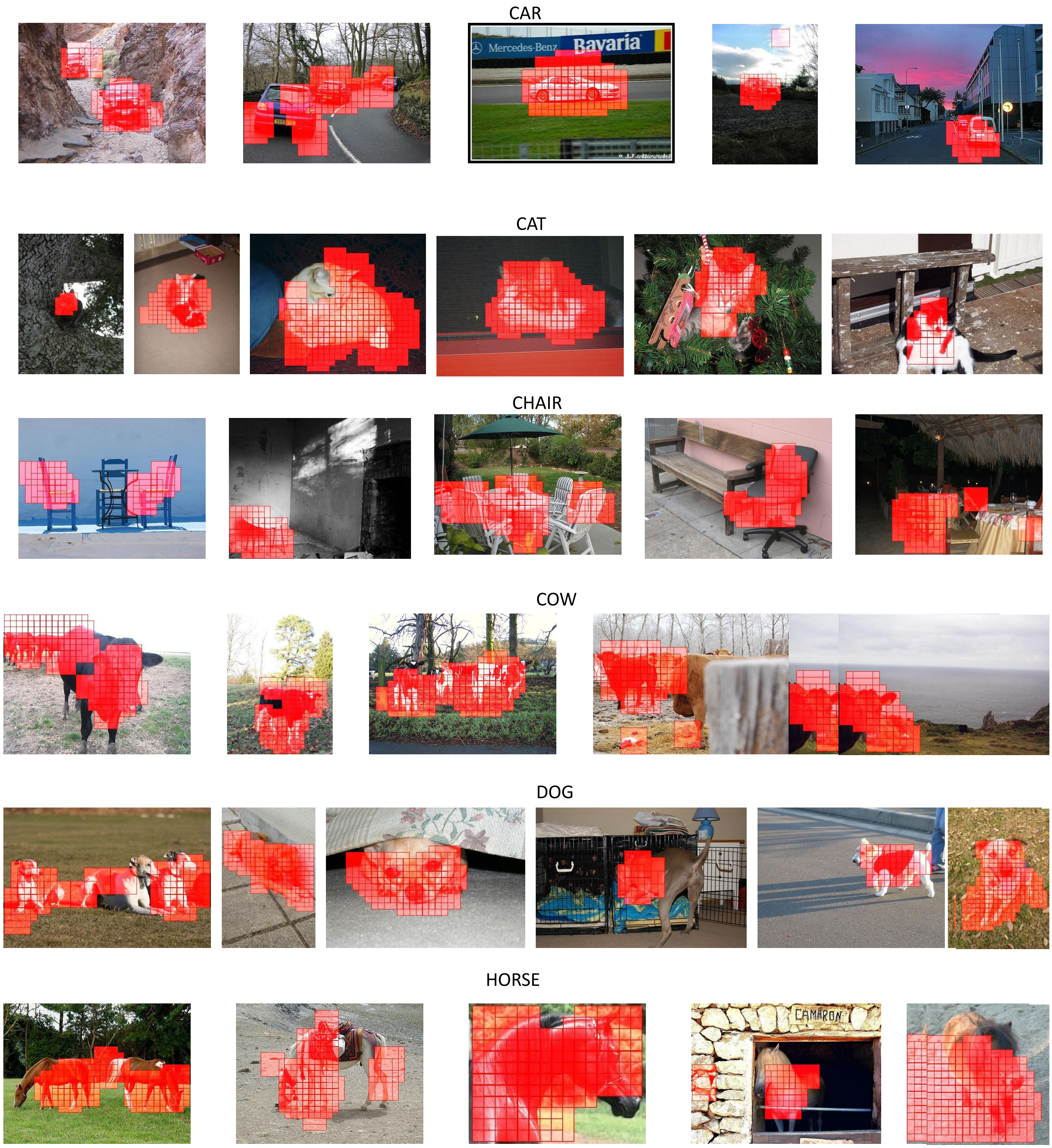


Figure 10. Class-specific patches identified on PASCAL VOC-07 images.

7. Supplementary results for applications

7.1. Object class segmentation

7.1.1 Comparison with co-segmentation approaches on Object Discovery dataset

Segmentation results obtained using our saliency maps are compared with OD (object discovery) [29], Joulin *et al.* [15], and DPM+Grabcut implementation given in [1]. We did not compare with semantic object selection [1] since their training requires an additional level of supervision to select training images having white background.

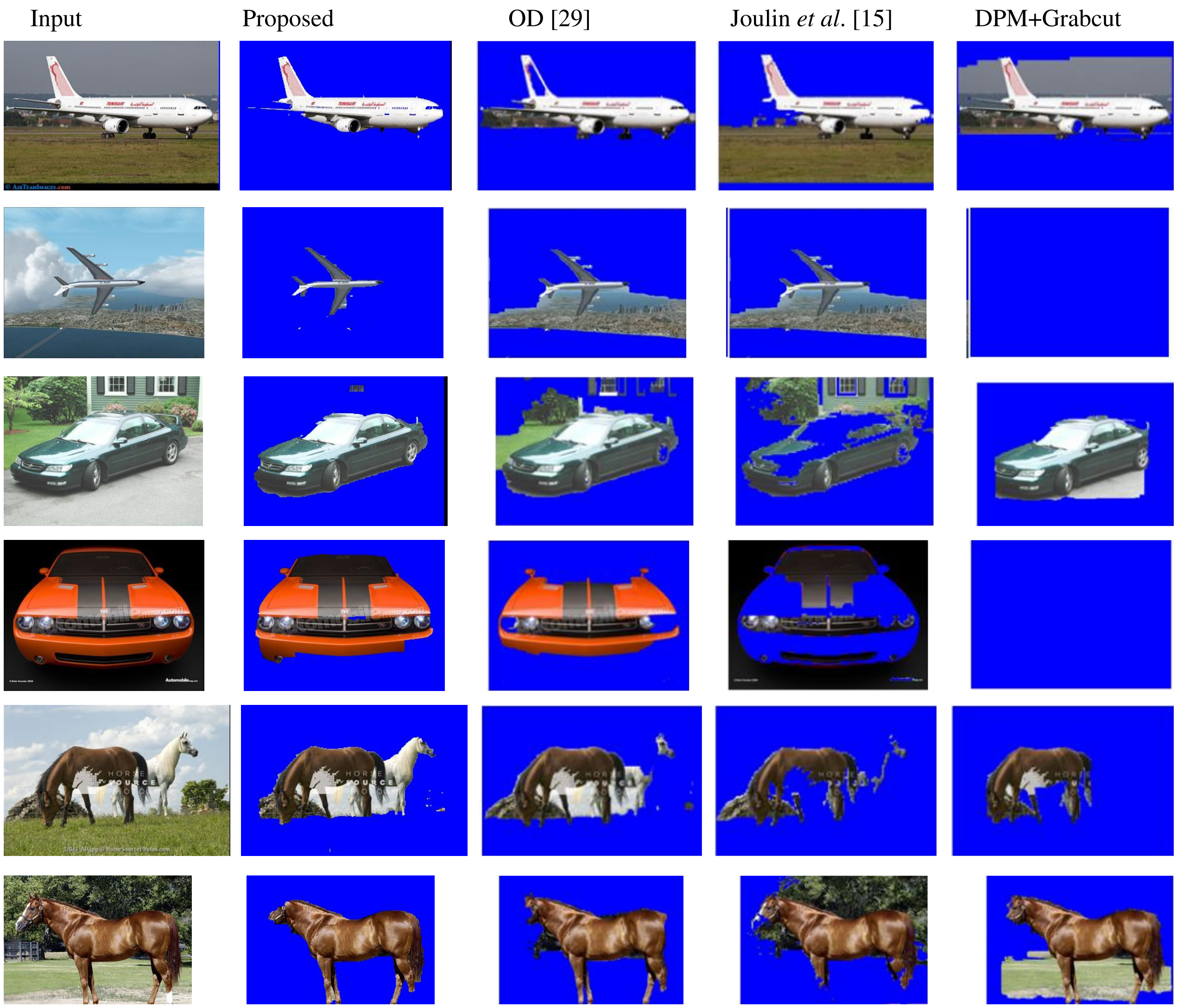


Figure 11. Comparison with co-segmentation approaches -object discovery [29], Joulin *et al*. [15] and grabcut applied on DPM detection output on Object Discovery dataset.

7.1.2 More segmentation results

Figure 12. Segmentation results obtained from our saliency maps on Object Discovery dataset.

7.2. Object annotation

Fig. 13 shows the object annotation results on PASCAL VOC-07 detection images. Annotaiton box obtained from our saliency map are shown in yellow color and the ground truth is shown as green colored box.

Figure 13. Object annotation obtained on PASCAL VOC-07 detection training images using proposed approach. Green rectangular boxes shows the ground truth and yellow boxes indicates the annotation boxes obtained using the proposed approach

7.3. Action-specific patch discovery

A linear classifier is learned for each action category in PASCAL VOC 2010 action dataset and these ScSPM-based classifiers are used to extract the category-specific patches corresponding to each action class as explained in the paper. Action classification works like [32] use a bounding box across the person even on test images. For training and testing, we do not use a bounding box across the person. A binary label indicating the presence or absence of a particular action in the given image is only used for classifier training. It can be seen from the results below that the proposed framework is able to extract class-specific patches corresponding to each action.

7.3.1 Action specific patches identified by R-ScSPM saliency

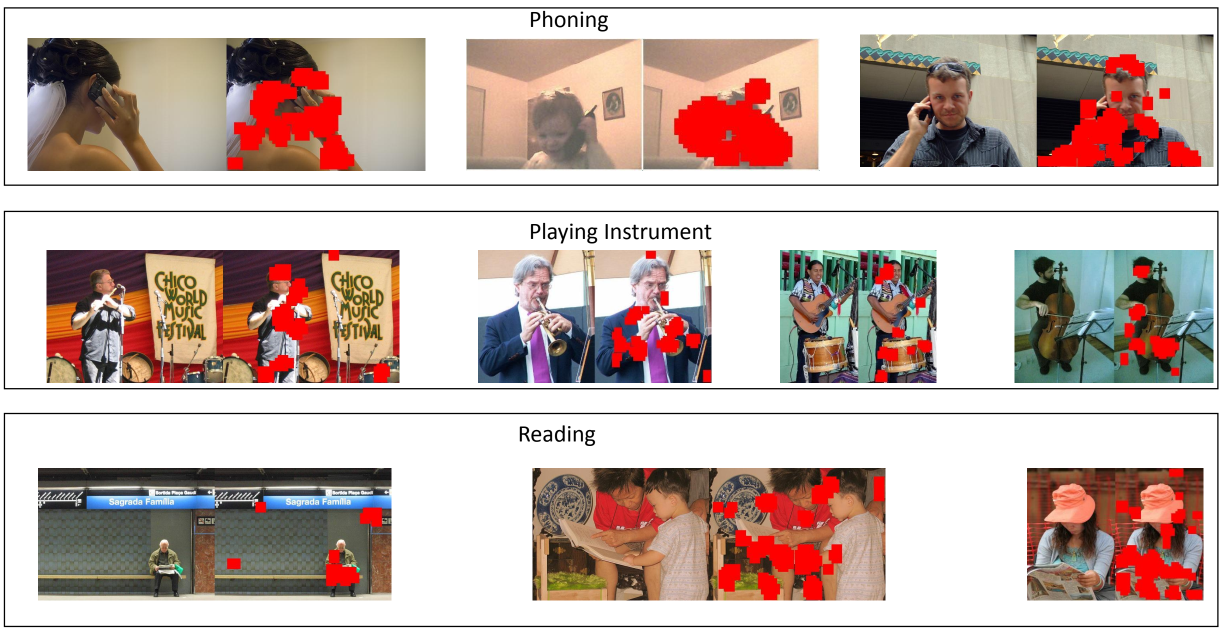


Figure 14. Action category-specific patches identified on PASCAL VOC 2010 action dataset training images by the proposed patch selection strategy; i.e, by thresholding R-ScSPM saliency at *0.5*). It can be observed that for *phoning* category, hand near the ear together with phone are identified as the class specific patches. The instruments in *playing instrument*, books and newspaper in *reading* are detected as category specific patches.

References

The numbering of references in the paper are used for citation.