WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation – Supplementary

Thibaut Durand[(1)⋆], Taylor Mordan[(1,2)⋆], Nicolas Thome[(3)], Matthieu Cord[(1)]
(1) Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, 4 place Jussieu, 75005 Paris
(2) Thales Optronique S.A.S., 2 Avenue Gay Lussac, 78990 Élancourt, France
(3) CEDRIC - Conservatoire National des Arts et Métiers, 292 rue St Martin, 75003 Paris, France
{thibaut.durand, taylor.mordan, nicolas.thome, matthieu.cord}@lip6.fr

1. Experimental setup for classification

In this section, we give detailed information about the explored datasets. Table 1 summarizes the number of images for training and testing, the number of classes and the evaluation measures. Table 2 gives information about our multiscale setup.

Dataset	Train	Test	Classes	Eval.
VOC07	5,011	4,952	20	MAP
VOC12	11,540	10,991	20	MAP
VOCAction	2,296	2,292	10	MAP
COCO	82,783	40,504	80	MAP
MIT67	5,360	1,340	67	accuracy
15 Scene	1,500	2,985	15	accuracy

Table 1. Dataset information: number of train and test images, number of classes and evaluation measures (MAP: Mean Average Precision).

Image size	Size before pooling	k^+, k^-
224×224	7×7	3
280×280	9×9	5
320×320	10×10	10
374×374	12×12	15
448×448	14×14	20
560×560	18×18	25
747×747	24×24	30

Table 2. Multiscale setup. We detail the input image sizes, along with the sizes of the feature maps before spatial pooling and the parameter values used in the spatial pooling.