Supplementary Material: Object Extent Pooling for Weakly Supervised Single-Shot Localization

Amogh Gudi\textsuperscript{1,2} \texttt{amogh@vicarvision.nl}
Nicolai van Rosmalen*\textsuperscript{1,2} \texttt{nicolai@vicarvision.nl}
Marco Loog\textsuperscript{2} \texttt{m.loog@tudelft.nl}
Jan van Gemert\textsuperscript{2} \texttt{j.c.vangemert@tudelft.nl}

\textsuperscript{1} Vicarious Perception Technologies
Amsterdam, The Netherlands
\textsuperscript{2} Delft University of Technology
Delft, The Netherlands

Abstract

This document contains some additional figures to supplement the contents of the paper ‘Object Extent Pooling for Weakly Supervised Single-Shot Localization’. All details about the figures are included in their captions.

- **Figure 1** demonstrated the forward and backward pass through the SPAM pooling layer for an example input/gradient.
- **Figure 2** is provided to highlight the differences in the backprojected areas between CAMs trained by global max pooling (GMP), global average pooling (GAP) and our spatial pyramid averaged max (SPAM) pooling methods.
- **Figure 3** shows bird localization examples on a weakly labelled dataset of CCTV images from a nature reserve.
- **Figures 4–23** show localization examples on images from the PASCAL VOC dataset (test set).

* Equal contribution as the first author.

© 2017. The copyright of this document resides with its authors. It may be distributed unchanged freely in print or electronic forms.
Figure 1: Illustration of SPAM’s forward and backward pass during training. During the forward pass, the CAM layer’s activations serve as the input to the SPAM pooling layer (bottom left). These activations pass through the pyramid of local average pooling and global max pooling as part of the layer’s forward pass. Note that the operations of the first and last pyramid steps with 1x1 and NxN average pooling kernels resemble that of global max pooling and global average pooling layers, respectively. Similarly, during the backward pass, that the gradients (shown in 3D) of the pyramid steps with 1x1 and NxN average pooling kernels are effectively the same as those of global max pooling and global average pooling, respectively.
Figure 2: **Effect of global pooling on backprojection.** This figure shows the training effect of the three global pooling types (global max, global average and global SPAM) on the backprojection of the CAM activations of a LeNet-5 based network. The network was trained on the MNIST128 dataset to classify digit 3 as the positive class. On this typical example image, it can be seen that the backprojection area of a GAP trained CAM is very large, while a CAM trained with GMP backprojects onto a too-small area, likely containing the most discriminative part of the object. The SPAM trained CAM’s backprojection more closely aligns with the true boundaries of the positive class object.

Figure 3: **Examples of bird localization** on a weakly labelled dataset of CCTV images from a nature reserve. The network used was a SPAM-pooling trained CAM network (VGG-16 based). The bottom row shows the magnified version of the localized birds in additional images.
Figure 4: **Aeroplane** class localization examples from the PASCAL VOC dataset (test set).

Figure 5: **Bicycle** (bike) class localization examples from the PASCAL VOC dataset (test set).

Figure 6: **Bird** class localization examples from the PASCAL VOC dataset (test set).

Figure 7: **Boat** class localization examples from the PASCAL VOC dataset (test set).

Figure 8: **Bottle** class localization examples from the PASCAL VOC dataset (test set).
Figure 9: **Bus** class localization examples from the PASCAL VOC dataset (test set).

Figure 10: **Car** class localization examples from the PASCAL VOC dataset (test set).

Figure 11: **Cat** class localization examples from the PASCAL VOC dataset (test set).

Figure 12: **Chair** class localization examples from the PASCAL VOC dataset (test set).

Figure 13: **Cow** class localization examples from the PASCAL VOC dataset (test set).
Figure 14: **Table** class localization examples from the PASCAL VOC dataset (test set).

Figure 15: **Dog** class localization examples from the PASCAL VOC dataset (test set).

Figure 16: **Horse** class localization examples from the PASCAL VOC dataset (test set).

Figure 17: **Motorcycle** class localization examples from the PASCAL VOC dataset (test set).

Figure 18: **Person** class localization examples from the PASCAL VOC dataset (test set).
Figure 19: **Potted Plant** class localization examples from the PASCAL VOC dataset (test set).

Figure 20: **Sheep** class localization examples from the PASCAL VOC dataset (test set).

Figure 21: **Sofa** class localization examples from the PASCAL VOC dataset (test set).

Figure 22: **Train** class localization examples from the PASCAL VOC dataset (test set).

Figure 23: **TV Monitor** class localization examples from the PASCAL VOC dataset (test set).