# Generalized orderless pooling performs implicit salient matching

Marcel Simon[1], Yang Gao[2], Trevor Darrell[2], Joachim Denzler[1], Erik Rodner[3]
[1] Computer Vision Group, University of Jena, Germany [2] EECS, UC Berkeley, USA
[3] Corporate Research and Technology, Carl Zeiss AG
{marcel.simon, joachim.denzler}@uni-jena.de {yg, trevor}@eecs.berkeley.edu

## A. Further examples for the classification visualization

In Sect. 4 of the paper, we visualize the classification decision for a single given test image using $\alpha$-pooled features. Figure 2 and 3 depict additional randomly selected examples of these visualizations for all datasets used in the paper, namely CUB200-2011 birds, Oxford flowers 102, Stanford cars, Caltech 256, MIT-67 scenes. Fine-tuning was used on all datasets.

## B. Quantification of the classification visualization versus $\alpha$

Figure 8 in the paper investigates the influence of $\alpha$ by giving classification visualizations for different values of $\alpha$. In this section of the supplementary material, we extend this analysis by giving a quantitative evaluation of the influence. In particular, we are interested in the peak contribution of training image regions to the classification decision.

Fig. 1 depicts the results for the CUB200-2011 birds and MIT-67 scenes dataset. For $\alpha \in \{1, 2, 3\}$, the box plot shows the distribution of the peak contributions. The contribution is measured in percent of all contributions amongst the top-10 most relevant training images for each test image. It can be clearly seen, that larger values of $\alpha$ dramatically increase the influence of the highest contributing region. For example, the median peak contribution almost doubles on MIT scenes, going from around 8.4% ($\alpha = 1$) up to 15.0% ($\alpha = 3$). In case of CUB200-2011 birds, the median increases from 13.4% ($\alpha = 1$) to 19.7% ($\alpha = 3$).

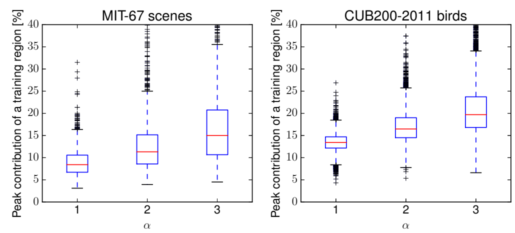


Figure 1. Peak contribution of a single training region to a classification decision as shown in Fig. 4 and 8 in the paper. For each test image, we calculate the classification visualization and compute statistics over the region with the largest contribution for each image. In the plot, we show the maximum contribution in % to the classification decision as shown in the top left of each image in Fig. 4 and 8.

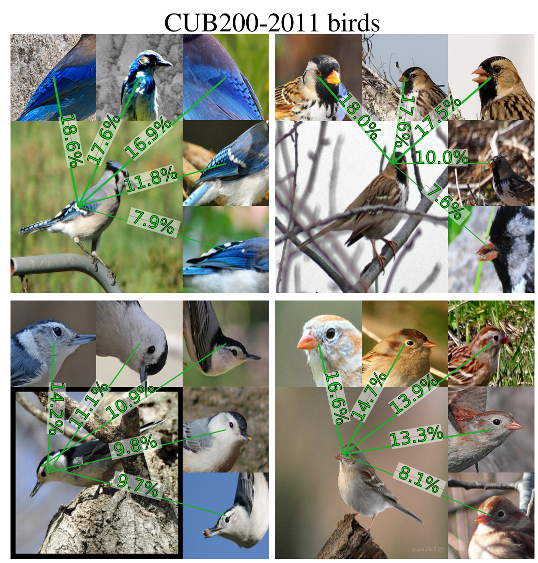


Figure 2. Visualization of the most relevant corresponding image regions for the classification decision: The large image in the bottom left corner is the test image and the surrounding images are crops of the training examples with highly relevant image regions. Correspondences are highlighted in green, if the training image belongs to the correct class of the test image, and red for false predictions. Percentages show the relative impact of the correspondence on the final decision score. Additional randomly selected samples continuing Fig. 2 of the paper for CUB200-2011.

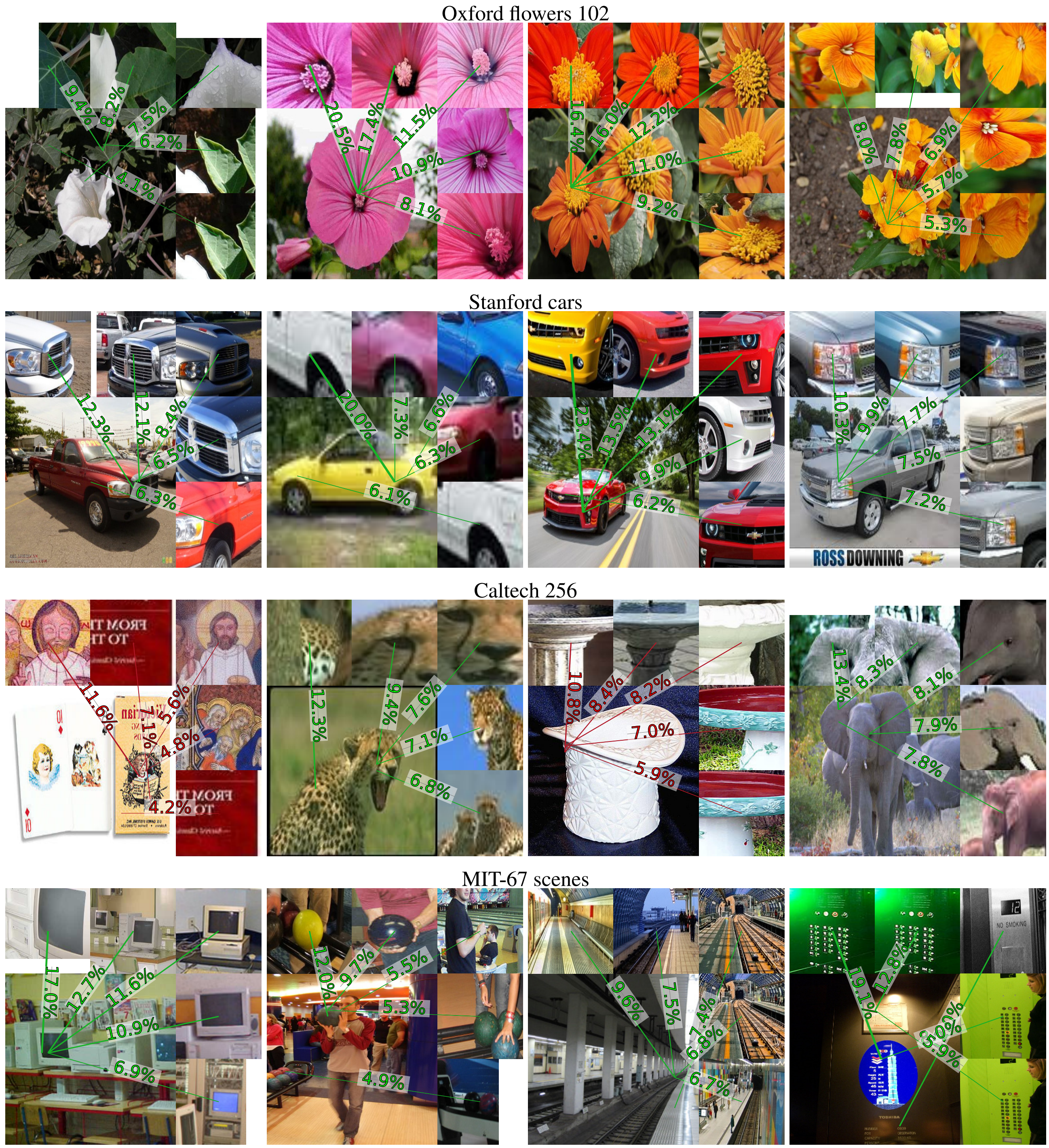


Figure 3. Visualization of the most relevant corresponding image regions for the classification decision (continued from Fig. 2). Sample visualization for the datasets Oxford flower 102, Stanford cars, Caltech 256 and MIT-67 scenes.