Supplementary Material - Compositional Convolutional Neural Networks: A Deep Architecture with Innate Robustness to Partial Occlusion

Adam Kortylewski Ju He Qing Liu Alan Yuille Johns Hopkins University Baltimore

A. Visualization of vMF Kernels



Figure 1: Illustration of vMF kernels μ_k by visualizing image patterns from the training data that activate a vMF kernel the most. Note that feature vectors that are similar to one of the vMF kernels, are often induced by image patches that are similar in appearance and often even share semantic meanings.

B. Visualization of Mixture Model Components



Figure 2: Visualization of mixture models $p(F|\theta_y^m)$ for M = 4 components. Each row show one of the classes: car, train, boat and bus. Note how images with different 3D viewpoint or different structure (boat) are approximately separated into different components.

C. Images used For Training of the Occluder Model



Figure 3: Natural images from which the parameters of the occluder model are learned. Note that no target object is present in any of these images.

D. Additional Qualitative Occluder Localization Results



Figure 4: Additional qualitative occlusion localization results on real images from the MS-COCO dataset. Each result consists of three images: The input image, the occlusion scores of a dictionary-based compositional model and the occlusion scores of our proposed CompositionalNet.

E. Qualitative Occluder Localization Results (no median filter)



Figure 5: Unfiltered occlusion localization results of Figure 5 in paper. Each result consists of three images: The input image, the occlusion scores of a dictionary-based compositional model and the occlusion scores of our proposed CompositionalNet. Note how our model can localize occluders with higher accuracy across different objects and occluder types for real as well as for artificial occlusion.