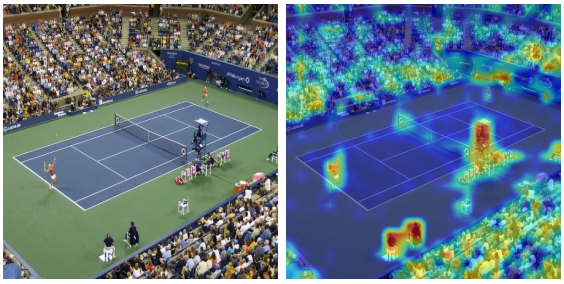
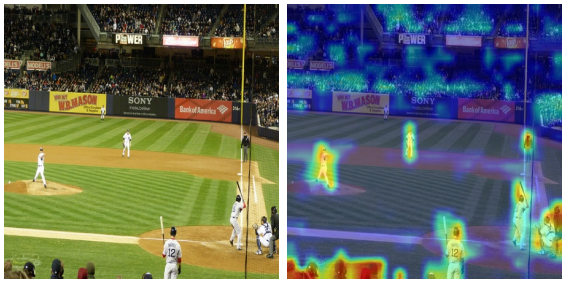
(a) 'person' class pair 1

(b) 'person' class pair 2

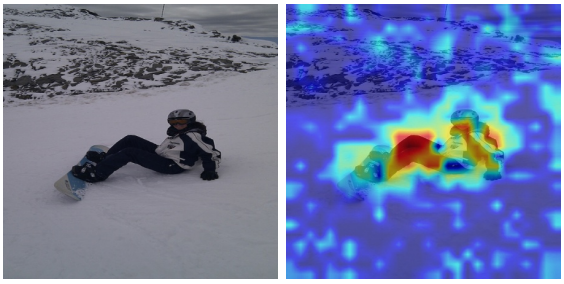
(c) 'person' class pair 3



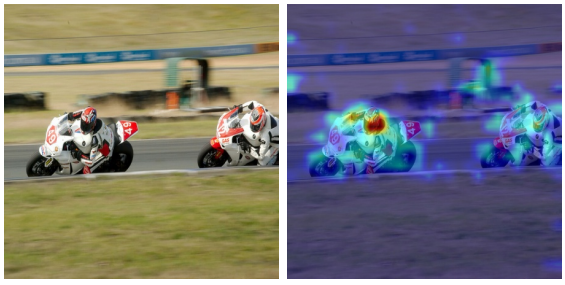
(d) 'person' class pair 4



(e) 'person' class pair 5



(f) 'person' class pair 6



(g) 'person' class pair 7

(h) 'person' class pair 8

Figure 1: Visualization of the 'person' class. There are in total 8 images sampled from MS-COCO 2014 (on the left of each pair) and the attention scores for the 'person' class overlapped on top of it (on the right). The score map was resized to the same size as that of the input image.

(a) ‘car’ class pair 1

(b) ‘car’ class pair 2

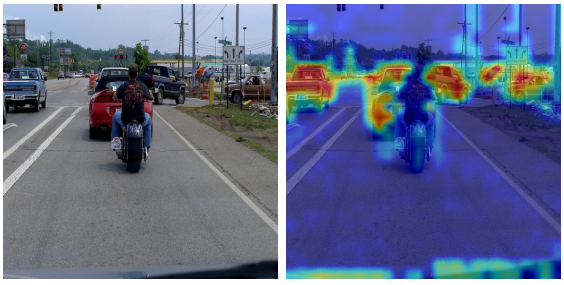
(c) ‘car’ class pair 3

(d) ‘car’ class pair 4

(e) ‘car’ class pair 5

(f) ‘car’ class pair 6

(g) ‘car’ class pair 7



(h) ‘car’ class pair 8

Figure 2: Visualization of the ‘car’ class. There are in total 8 images sampled from MS-COCO 2014 (on the left of each pair) and the attention scores for the ‘car’ class overlapped on top of it (on the right). The score map was resized to the same size as that of the input image.