The model is evaluated using a validation dataset of label information containing the exact location and class type of every object in the image. The model estimates the objects in the image and if the estimation contains at least the ground-truth number of each object, then the image is labelled as correct for each object class that was contained in the image. The results were then aggregated to the following values:

0 (Dalek): 39/45 = 86.67%

1 (cats): 41/47 = 87.23%

2 (dog): 49/55 = 89.09%

3 (person): 29/39 = 74.36%

4 (Lightsaber): 15/17 = 88.24%

5 (Red Lightsaber): 1/2 = 50.00%