Supplemental materials for Sketch2Photo

1 Intermediate results for scene item image selection

When counting false positive rates, only images of desired scene items with specified pose (i.e. contour) are considered as correct. All the other images are considered as false positive. For example, in Figure 4, only images with a red car viewed from the side are correct images.

Figure 2 is the image filtering with the keyword 'dog'. This example is to justify the search with verbal, e.g. 'dog jump'. Most of the dog images in Figure 2 do not have a dog in jumping pose. As a result, after all the filtering, the false positive rate is as high as 68%. In comparison, if it is searched with the keyword 'dog jump', this false positive ratio can be reduced to 15% by our filtering, as indicated in Table 1 in the paper. For each scene item, we manually count false positive rate of the top 100 ranked images in 3000 web search results.

Item:	Sailboat	Dog	Moto Rider	Red car	Wedding Kiss	Seagull	Sheep	Kid Ski	Salmon Jump	Bear Catch
IS(%):	71	97	86	87	77	72	73	70	73	93
SF(%):	61	97	81	83	78	70	69	67	71	94
CF1(%):	35	77	27	21	28	29	31	24	32	45
CF2(%):	27	68	24	19	19	23	20	21	22	33



Figure 1: Sailboat.



Figure 2: Dog.



Figure 3: Moto rider.



Figure 4: Red car.



Figure 5: Wedding kiss.



Figure 6: Seagull.



Figure 7: Sheep.



Figure 8: Kid ski.



Figure 9: Salmon jump.



Figure 10: Bear catch.

2 Composition Comparison



Figure 11: From top to bottom, each rows are blended by alpha blending [Wang and Cohen 2007], drag and drop pasting [Jia et al. 2006], photo clip art [Lalonde et al. 2007], and our hybrid method. The image parts are all from Figure 9 of our paper.



Figure 12: From left to right, each columns are input images, blended by alpha blending [Wang and Cohen 2007], drag and drop pasting [Jia et al. 2006], photo clip art [Lalonde et al. 2007], and our hybrid method. This figure demonstrates that our method has the ability to handle different illumination and complicated object boundary.

3 Automatic Compositions and User Refinement



Figure 13: (a) and (c) are the automatic compositions ranked top 10 in our system. (b) and (d) are compositions after user refinements. We highlight compositions with incorrect scene items with a blue frame. The three compositions chosen for the paper are highlighted by a red frame.



Figure 14: (a) and (c) are the automatic compositions ranked top 10 in our system. (b) and (d) are compositions after user refinements. We highlight compositions with incorrect scene items with a blue frame. The three compositions chosen for the paper are highlighted by a red frame.