Where and Who? Automatic Semantic-Aware Person Composition
Supplementary Material

Fuwen Tan *[1], Crispin Bernier [1], Benjamin Cohen [1], Vicente Ordonez [1], and Connelly Barnes [1,2]

[1]University of Virginia, [2]Adobe Research

1. Implementation details of the prediction model

The proposed prediction network is implemented in Keras [1]. Most of the layers we exploit in our network have corresponding implementations in Keras except for the ROI slicing operation. During training, the inputs of the network are the inpainted and blurred color images as well as their corresponding rendered layouts. During testing, the inputs are the blurred background images and the rendered layouts. These images are first transformed into square images by padding, and then resized to a resolution of $(480, 480, 3)$. Before feeding the images into the network, the mean pixel color for the ImageNet dataset (in sRGB space, $(103.939, 116.779, 123.68)$) is subtracted from the image. The loss of the model is a weighted sum of the categorical cross entropy losses for the location and size prediction branches, with weights 1.0 and 2.0 respectively. To train the network, we use the Adam solver [2] with a fixed learning rate of 0.0001. We restrict the training data to the training split of MS-COCO dataset and use horizontal flipping for data augmentation. Our final model was trained for four epochs.

2. Additional qualitative results

Figure 1 and 2 shows additional heatmaps of predicted locations. Despite being trained to predict a fixed unique location, our network can still mimic the location distribution reasonably well.

Figure 3, 4 and 5 show additional top 1 composites from our system covering various scenes.

3. Demo video of the prototype user interface

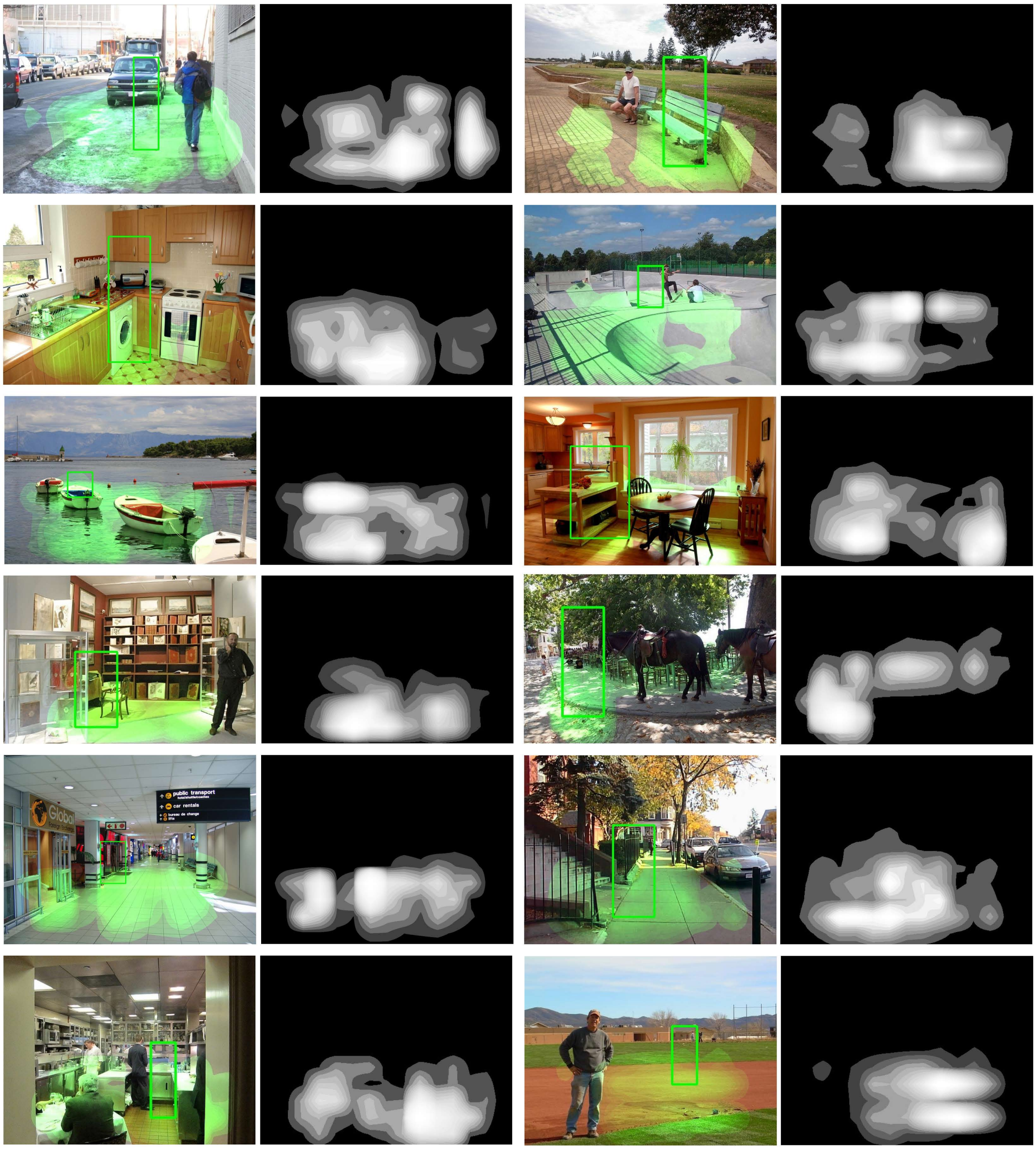
To demonstrate the potential application of our system, we developed a proof-of-concept user interface for composite image generation and interactive layout refinement. We include a supplementary video showing examples of the interactive editing.

*fuwen.tan@virginia.edu

References

[1] F. Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

[2] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.



(A) (B) (C) (D)

Figure 1: Additional heatmaps of predicted locations. The green boxes show top 1 bounding boxes from our system. **The images are manually selected from the test set.**

Figure 2: Additional heatmaps of predicted locations. The green boxes show top 1 bounding boxes from our system. **The images are randomly selected from 270 background images.**

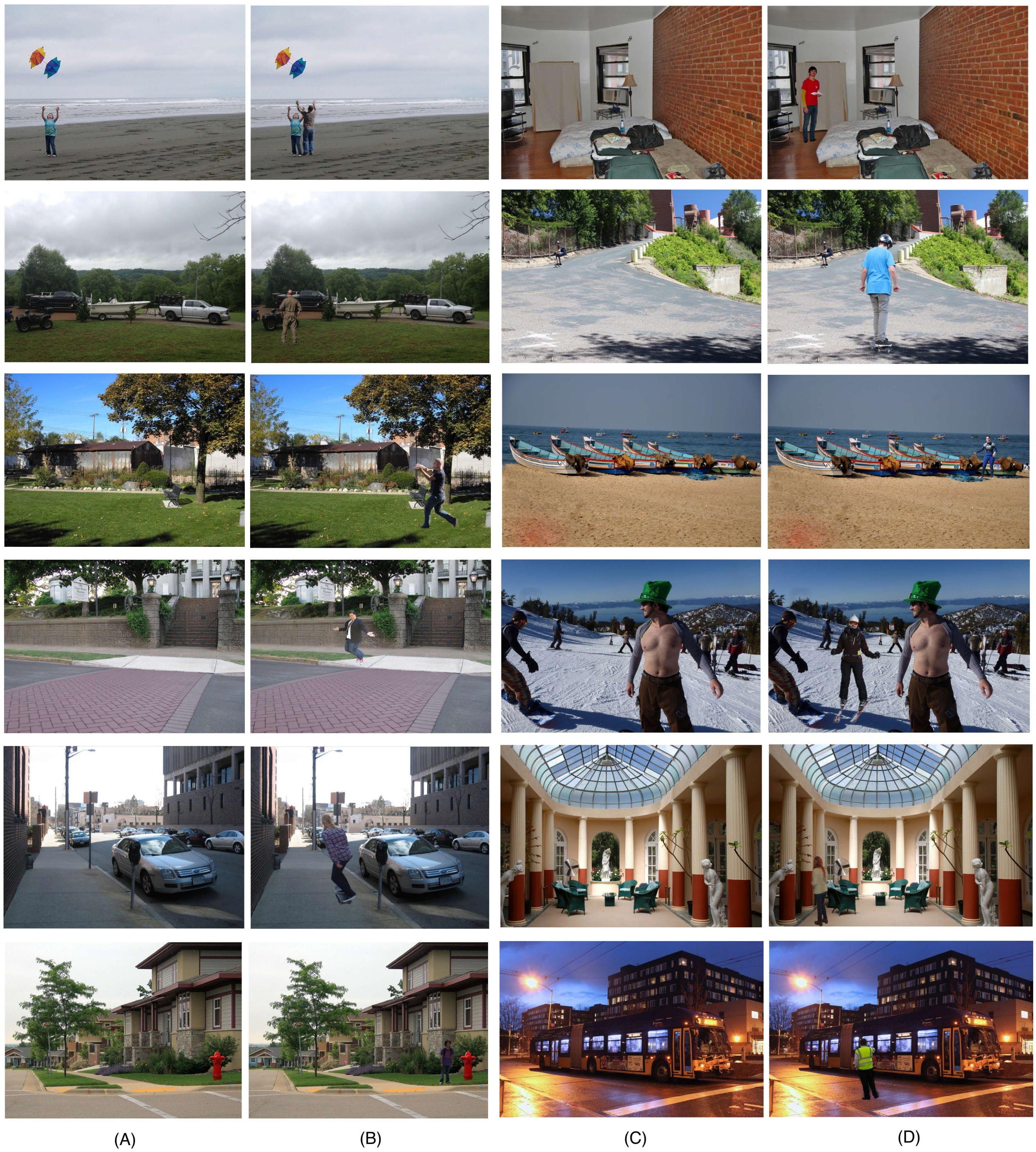


Figure 3: Additional top 1 composites automatically generated from our system. (A)(C): input images, (B)(D): composite results. **The images are manually selected from the test set.**

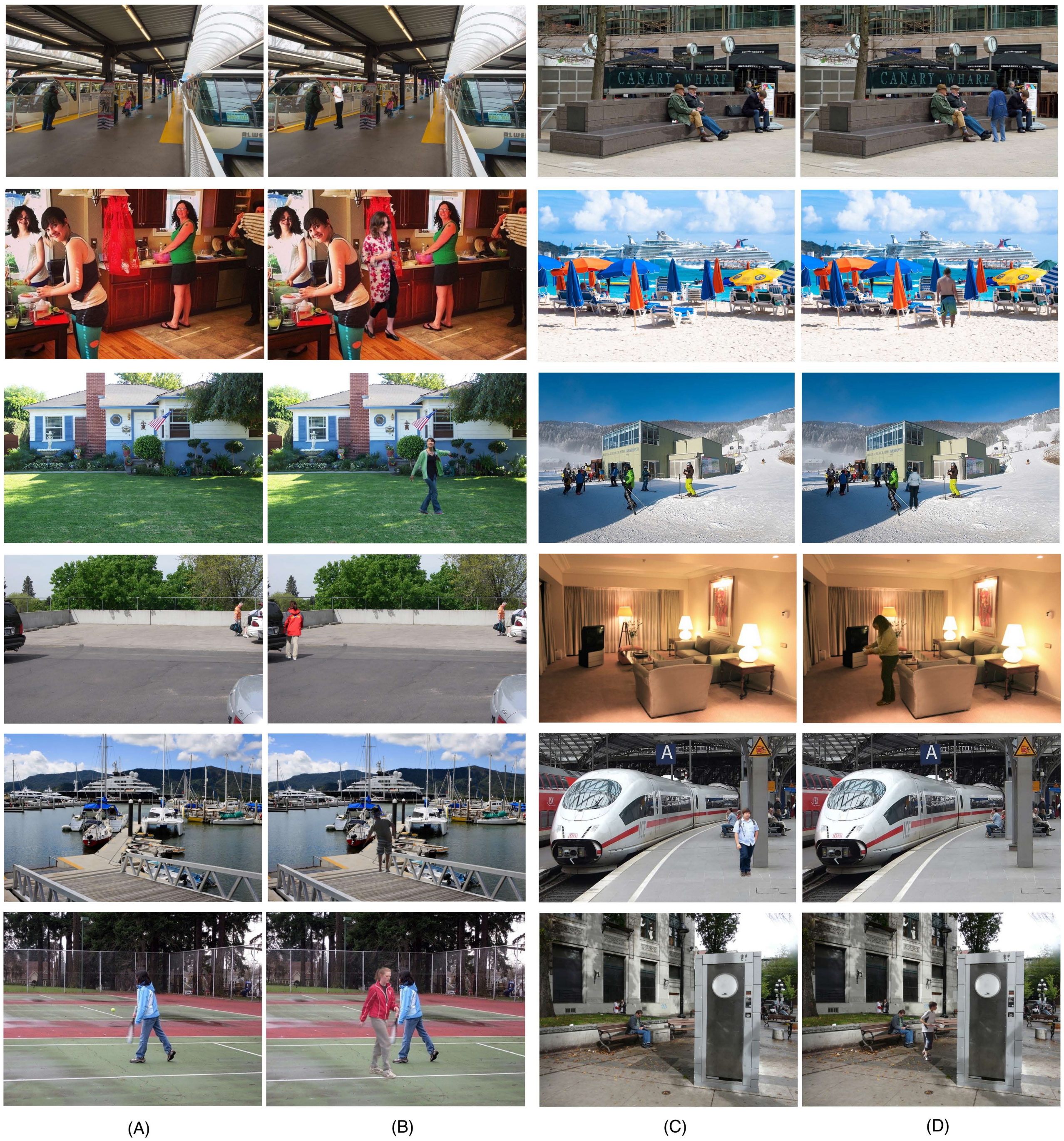


Figure 4: Additional top 1 composites automatically generated from our system. (A)(C): input images, (B)(D): composite results. **The images are manually selected from the test set.**

Figure 5: Additional top 1 composites automatically generated from our system. (A)(C): input images, (B)(D): composite results. **The images are randomly selected from the test set.**