Supplementary materials for
“Reasoning-RCNN: Unifying Adaptive Global Reasoning into Large-scale Object Detection”

Paper ID: 3864

Improvement on infrequent categories

We calculate the incremental overall AP for top 200 infrequent categories from Faster-RCNN to Resonning-RCNN with relation knowledge. The bar chart shows results in Figure 1. Solid improvement for our method for those categories can be found. This speaks well for our method can boost the performance of rare categories. Note that some of the categories are degraded . We checked some of them and find that those categories has few samples in VG. Their annotations of relationship are biased so that the drop in performance of those categories may due to the noisy data.

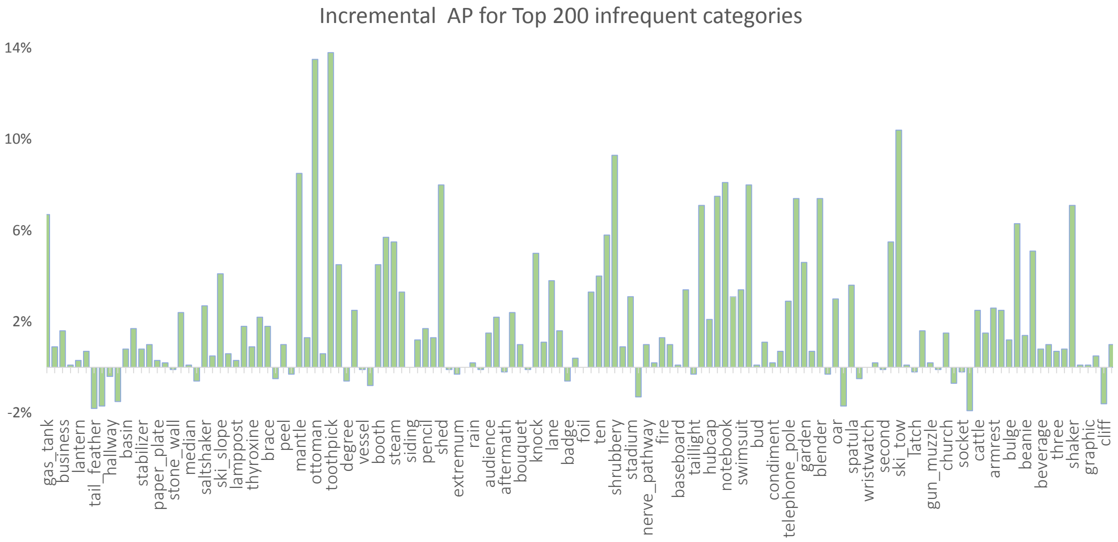


Figure 1: Incremental AP for top 200 infrequent categories from Faster-RCNN to Resonning-RCNN.

Visualization for parts of the relation and attribute knowledge graphs

We visualize relation and attribute knowledge graphs in Figure 2. The $C \times C$ knowledge graphs are captured from frequency of VG_{3000} annotation with symmetric transformation and row normalization. In the relation knowledge graph, the pair of “knife” and “spoon” are more likely to appear at the same time, and the co-occurrence relationship between “orange” and “apple” is higher than “orange” and “sandwich”. In the attribute knowledge graph, tablewares have more same attributes with each other and “broccoli” is similar with hot_dog in color and usage.

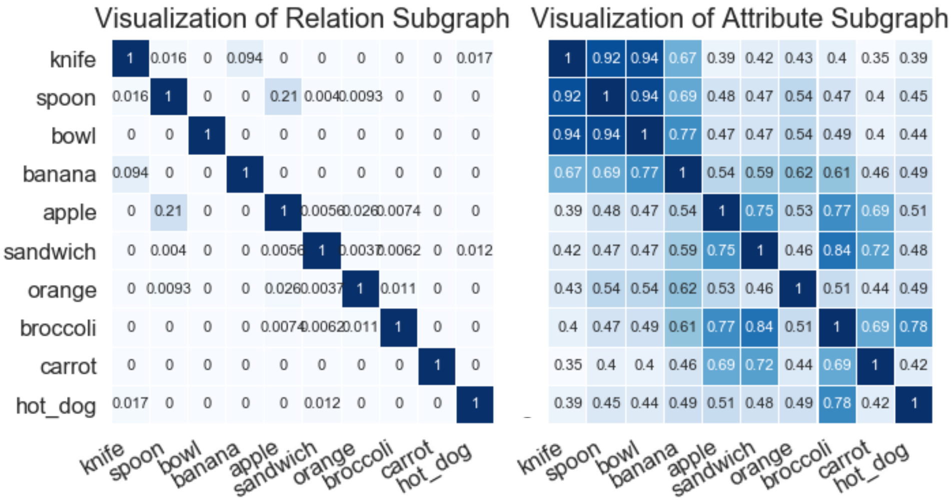
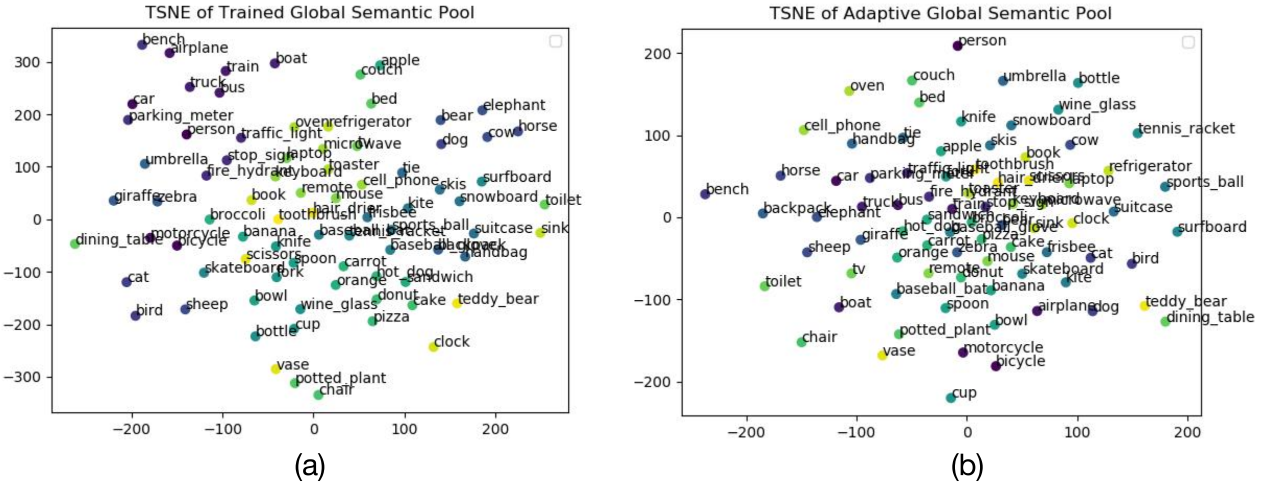


Figure 2: Visualization for parts of the relation and attribute knowledge graphs.

Analysis of adaptive attention interpretability

To show benefit of adaptive attention mechanism, we visualize Global Semantic Pool during testing in Figure 3. In the inference phrase, Global Semantic Pool collected from previous classifier is frozen and scattered. While (b) shows that the adaptive attention module can cluster some important categories in the Global Semantic Pool according to the context of the image.



(a)

(b)

Figure 3: 2D visualization of Global Semantic Pool by t-SNE method.

Gradient visualization for the Resonning-RCNN with relation knowledge

To further investigate the effectiveness of Resonning-RCNN, we make use of some gradient visualization technique to visualize the detection results from a "deconvolution approach". Specifically, we adopt the gradient visualization with guided back-propagation (Springenberg et al., 2015) on the final layer of the Faster-RCNN and Resonning-RCNN with relation knowledge. Figure 4 shows three representative examples of gradient visualization comparing Faster-RCNN with Resonning-RCNN. Left panels denote the original images and detection results with a visualization threshold of 0.5, the middle panels denote colored guided back-propagation graphs and the guided back-propagation saliency graphs.

From Figure 4, we can see the activation of our method on the middle and right panels is clearer than the Faster-RCNN. For example, from the top comparison, we can find that more areas in the refrigerator are activated. Thus more objects such as bag, shelf and glass can be located and recognized for our method. For the middle comparison, the activation of objects around baby and man are more obvious. This may due to our method catches the relationship between people and its surrounding objects thus these areas are activated. From the last comparison, we can see a clear train in our method's saliency graph since the Resonning-RCNN considers the relationship on the train such as window and light. That's why our Resonning-RCNN can detect more object on the train.

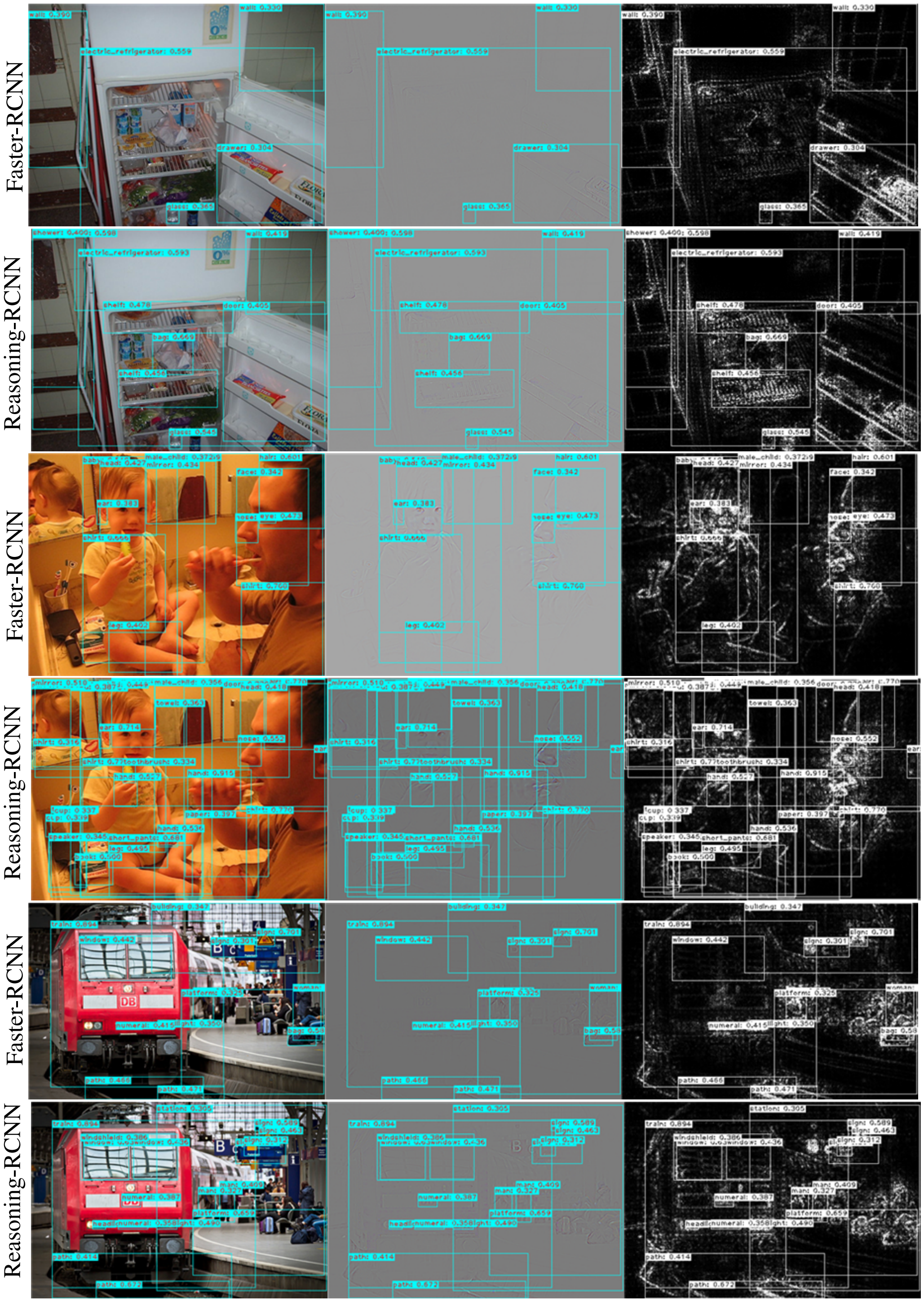


Figure 4: Gradient visualization with guided back-propagation from the output layer of Faster-RCNN and our Reasoning-RCNN.

Additional Qualitative results

More qualitative comparisons on VG_{1000} between Faster-RCNN and our Reasoning-RCNN can be found in Figure 5. From the comparisons, objects with occlusion, ambiguities and rare category can be detected and localized well by our method, while the Faster-RCNN fails to detect them.

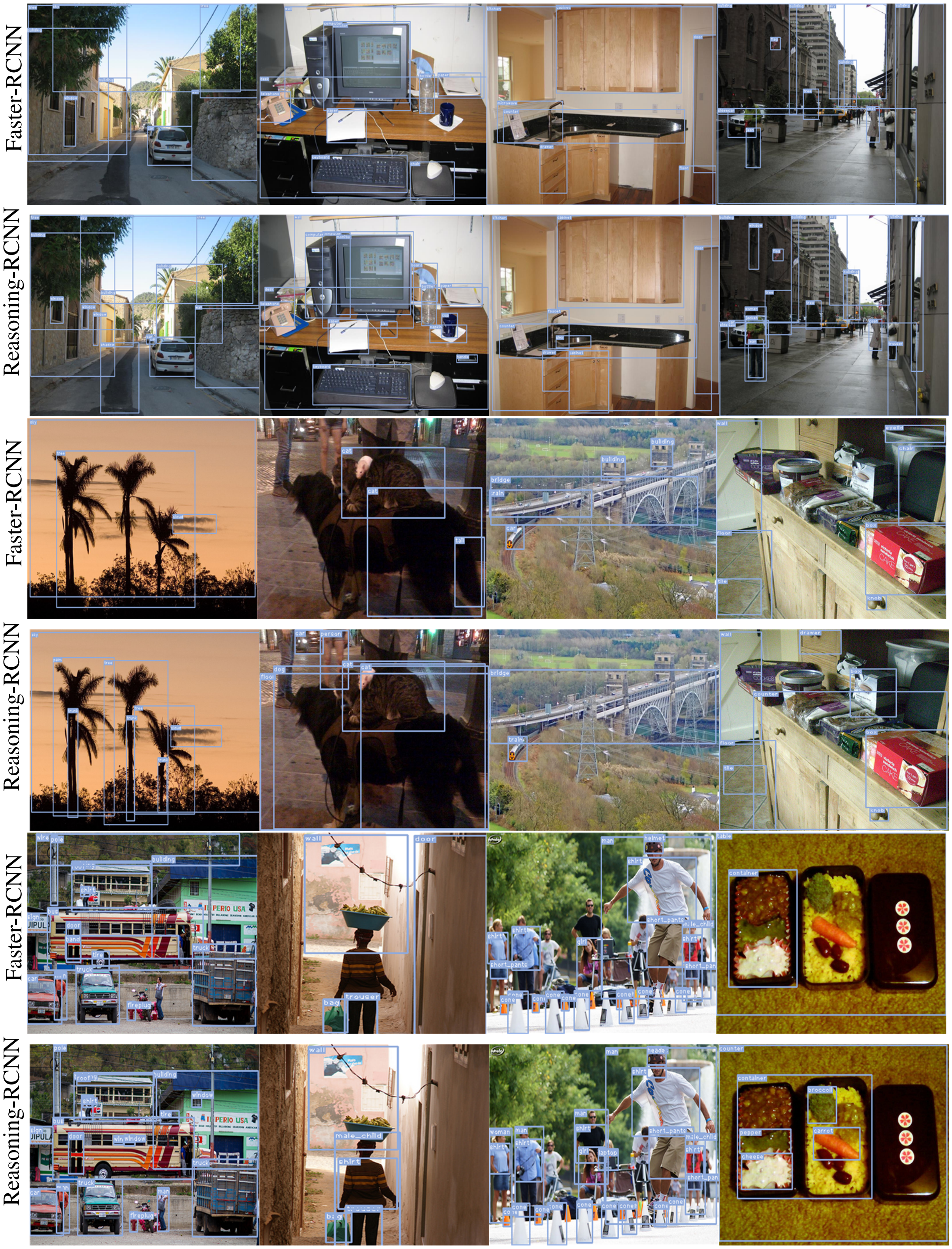


Figure 5: More qualitative results comparison on VG_{1000} between Faster R-CNN and Reasoning-RCNN. Objects with occlusion, ambiguities and rare category can be detected by our method.

References

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2015). Striving for simplicity: The all convolutional net. In *ICLR Workshop*.