A. Supplementary Material: Unconditional Scene Graph Generation

This document supplements our paper *Unconditional Scene Graph Generation* with dataset-level statistics, the mathematical description of the MMD kernels, and additional results on the different applications.

A.1. Dataset-level statistics

In addition to the MMD metrics (sample level comparison), we present the statistics to compare the generated and test samples on the dataset level. We compare 20k samples of generated scene graphs from SceneGraphGen, against the ground truth *i.e*. the test dataset from Visual Genome. Figure 1 reports different patterns such as object occurrence (a), relationship occurrence (b) and object co-occurrence (c). Object occurrence computes the occurrence probability of each object label over the whole dataset. Similarly, relationship occurrence computes the occurrence probability of each relationship label over the whole dataset. The object co-occurrence is computed by collecting the frequency (normalized) with which two *different* object labels co-occur in the same scene graph (scene). For improved visibility of co-occurrence patterns, we use a maximum threshold of 0.05. In all measurements in Figure 1, we observe similar patterns between the generated and ground truth datasets. Since each object category can occur multiple times in an image (instances), we compare the count distribution (1, 2, so on) for each object category using Kullback-Leibler (KL) divergence of generated dataset from the test dataset. Figure 2 shows the KL divergence of object count distribution for each object category, with a low average value of 0.048.

A.2. Object occurrences in the generated images

Figure 3 shows the object occurrence of detected objects (using FasterRCNN) in the images generated by StyleGAN (unconditional) vs the images generated by sg2im on scene graphs generated by SceneGraphGen (sg2im-SGG). We observe that sg2im-SGG generates images with more objects detected and better object statistics than StyleGAN. The FID of StyleGAN on Visual Genome is however 66.3, so better than sg2im-SGG, which we attribute to the respective image generative models, instead of the quality of input scene graph.

A.3. Additional examples from applications

We provide some additional examples for image generation in Figure 4, and scene graph completion in Figure 5.

A.4. Mathematical formulation of the MMD kernels

Here we give the details on the kernels used for the MMD evaluation.

A.4.1 Random walk graph kernel

A general formulation for random walk kernel is adopted from [2], which allows freedom to choose suitable node and edge kernels. In two graphs G_a and G_b, we want to compare two nodes r and s respectively. We can compare these two nodes by comparing all walks of length p in G_a starting from r against all walks of length p in G_b starting from s. The similarity between each walk-pair can be performed by comparing the respective nodes and edges encountered in the walks using suitable kernels. The kernel to compare any two nodes is given by

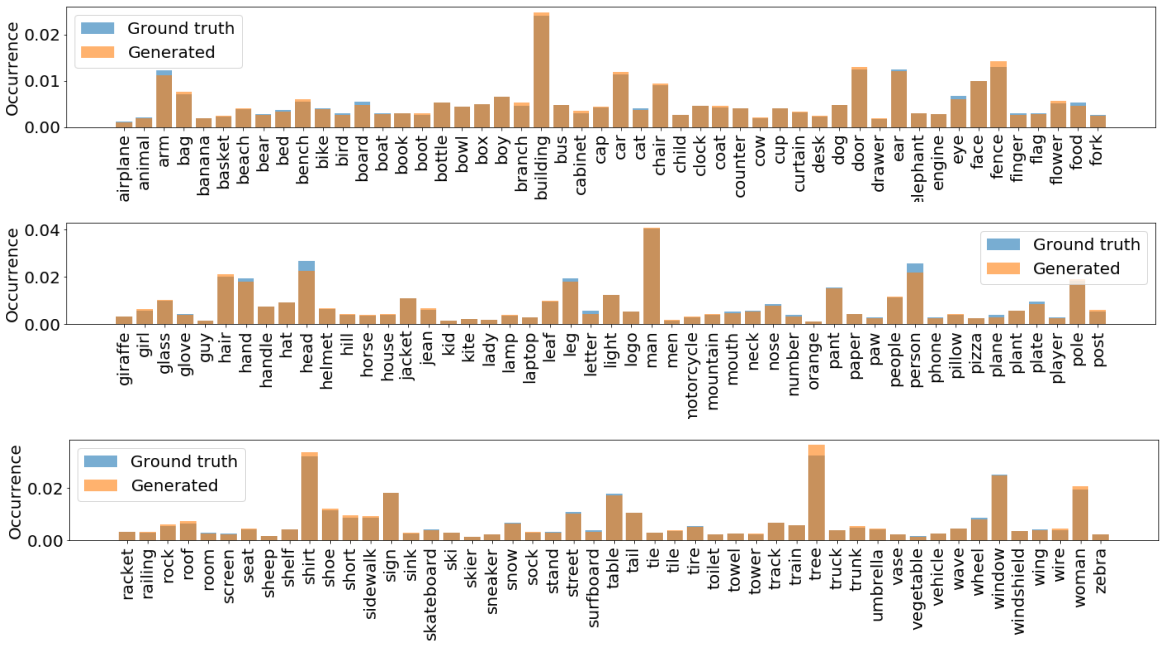
$$k_R^p(G_a, G_b, r, s) = \sum_{\substack{(r_1, e_1, .., e_{p-1}, r_p) \in W_{G_a}^p(r) \\ (s_1, f_1, .., s_{p-1}, f_p) \in W_{G_b}^p(s)}} \Big[k_{node}(r_p, s_p) \prod_{i=1}^{p-1} k_{node}(r_i, s_i) k_{edge}(e_i, f_i) \Big] \tag{1}$$

To compare the overall structure, the kernel in Equation 1 is summed over all pairs of nodes, and normalized with maximum of the kernel evaluation of each graph and itself.

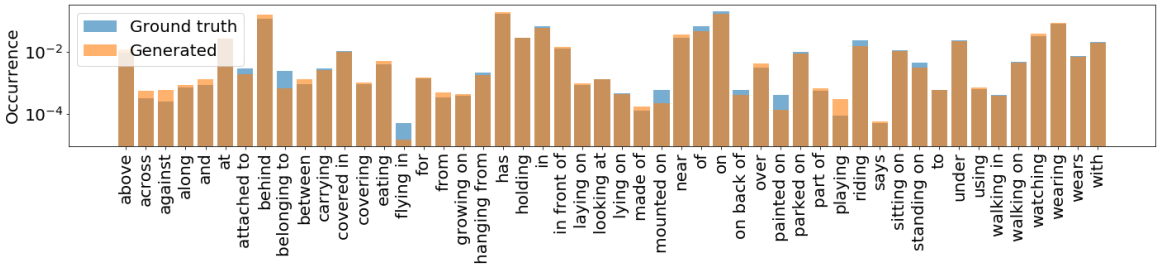
$$k_G^p(G_a, G_b) = \sum_{\substack{r \in V_{G_a} \\ s \in V_{G_b}}} k_R^p(G_a, G_b, r, s) \tag{2}$$

$$k_G^N(G_a, G_b) = \frac{k_G(G_a, G_b)}{\max(k_G(G_a, G_a), k_G(G_b, G_b))} \tag{3}$$

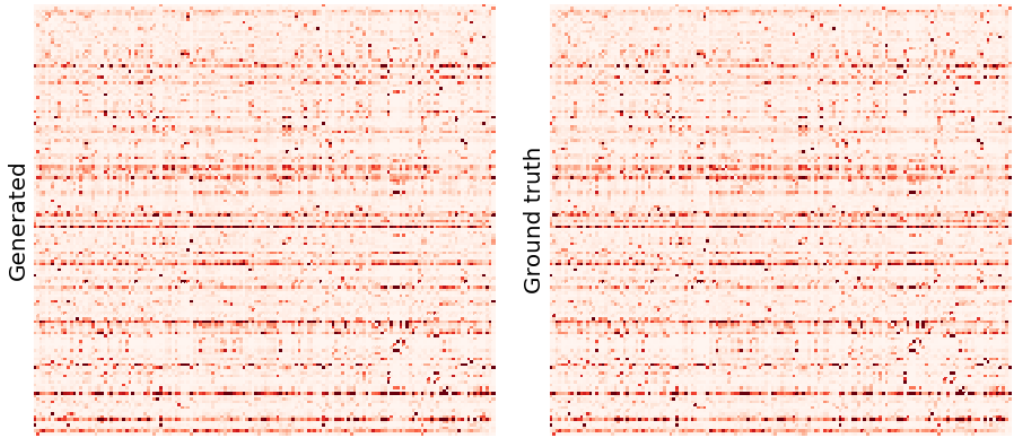
For comparing nodes, we use the simple Kronecker delta function which is 1 when the node categories match and 0 otherwise, *i.e*. $k_{node}(r, s) = \delta(r, s)$. However, since there are multiple nodes with the same category, the importance of the nodes in a graph will be lower for the category with one occurrence and higher for multiple occurrences. In fact, the importance of a category with multiple occurrences should diminish as the occurrences increase. For this purpose, the node kernel is normalized with the frequency of



(a) Object occurrence



(b) Relationship occurrence



(c) Object Co-Occurrence

Figure 1. Comparison of the dataset-level statistics of generated scene graphs against ground truth scene graphs from Visual Genome. a.) Object Occurrence, b.) Relationship Occurrence, c.) Object Co-Occurrence

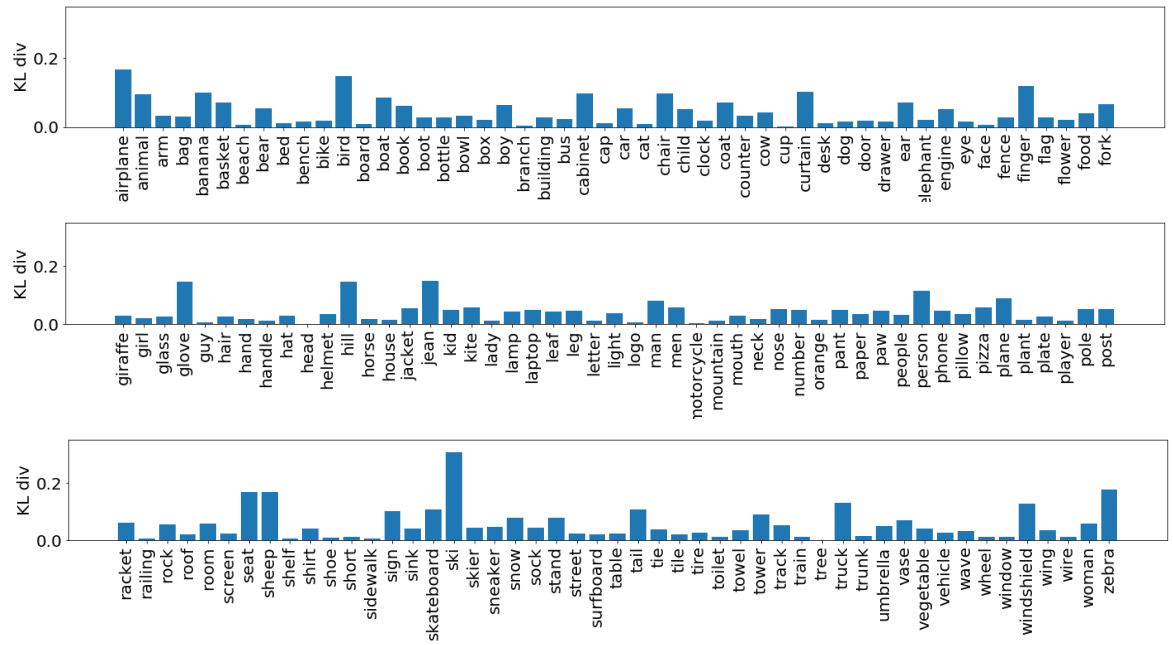


Figure 2. KL divergence of the generated dataset from test dataset, which compares the object count distribution (number of instances of a particular object category per scene) for each object category

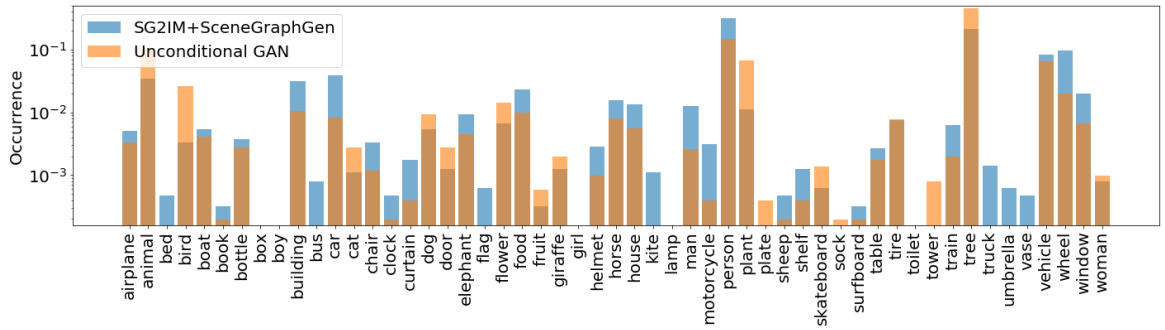


Figure 3. Comparison of occurrence of objects detected by Faster R-CNN on the images generated by Unconditional-GAN (StyleGAN2) vs. sg2im+SceneGraphGen model

occurrence in a graph. The node kernel is given by

$$k^N_{node}(r,s) = \sigma(r)\sigma(s)k_{node}(r,s),$$
$$\text{where} \quad \sigma(s) = \frac{1}{\sum_{s\in V_{G_s}} k_{node}(r,s)} \tag{4}$$

For comparing edges, we use the Kronecker delta function, *i.e.* $k_{edge}(p,q) = \delta(p,q)$.

A.4.2 Object set kernel

We want to compare two *sets* of object instances. Hein *et al.* [3] showed that for a domain set $\mathcal{X}$, two sets $A \in \mathcal{X}$, $B \in \mathcal{X}$, a positive definite kernels k_{label} and k_{count}, we can define a general set kernel between A and B as:

$$k_{set}(A,B) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{X}} k_{label}(x,y)k_{count}(A(x),B(y)) \tag{5}$$

We choose $k_{label}(x,y) = \delta(x,y)$ as the Kronecker delta function which is one when both x and y have the same object categories. $A(x)$ is the number of times element x appears in A and $B(y)$ is the number of times element y appears in B. k_{count} is defined as

$$k_{count}(A(x),B(y)) = \frac{1}{1+|A(x)-B(y)|} \tag{6}$$

k_{count} is the generalized t-student kernel [1, 4]. This formulation allows us to capture when the two sets have the

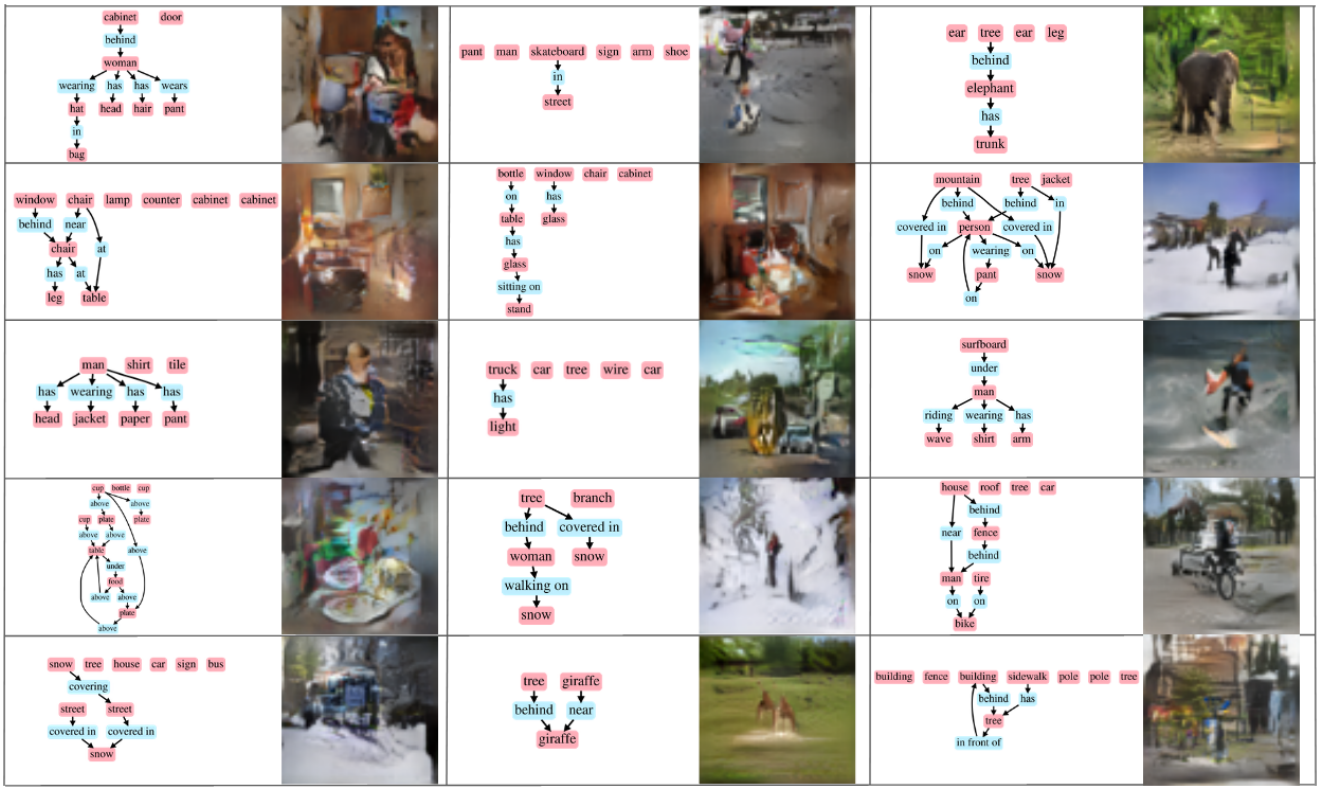


Figure 4. Additional examples of 64×64 images generated using sg2im on the corresponding scene graphs generated by SceneGraphGen

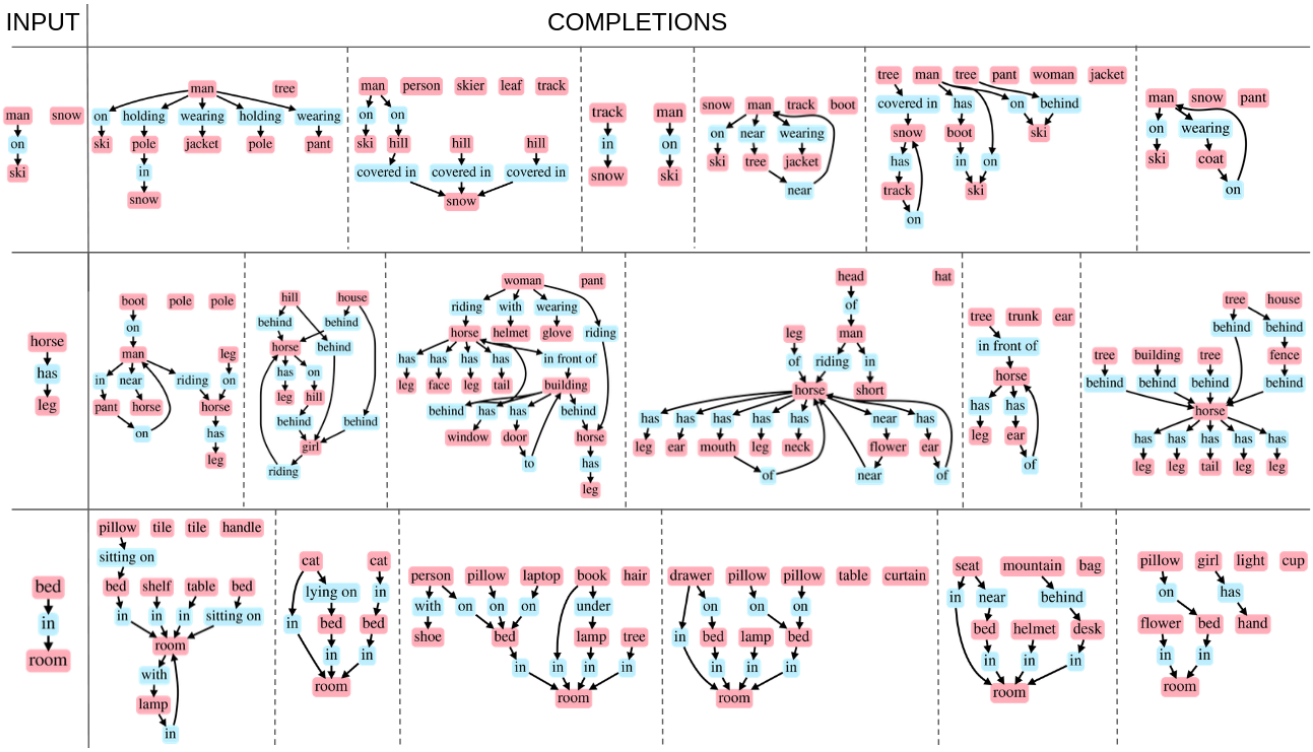


Figure 5. Additional examples of scene graph completions from an initial partial scene graph using SceneGraphGen

same object category member as well as how similar are the counts of those object category members. Similar to the graph kernel above, the object-set kernel is also normalized with the maximum of kernel evaluation of each object set with itself.

$$k^N_{set}(A,B) = \frac{k_{set}(A,B)}{\max(k_{set}(A,A), k_{set}(B,B))} \quad (7)$$

A.5. Anomaly detection comparison

Figure 6 demonstrates a comparison between SceneGraphGen and the GraphRNN baseline on the NLL plot, extending Figure 6 (left) in the main paper. Our interpretation is that our model is more sensitive to the level of dataset corruption compared to the baseline, *i.e.* the NLL gap between the different levels is larger than for the baseline.

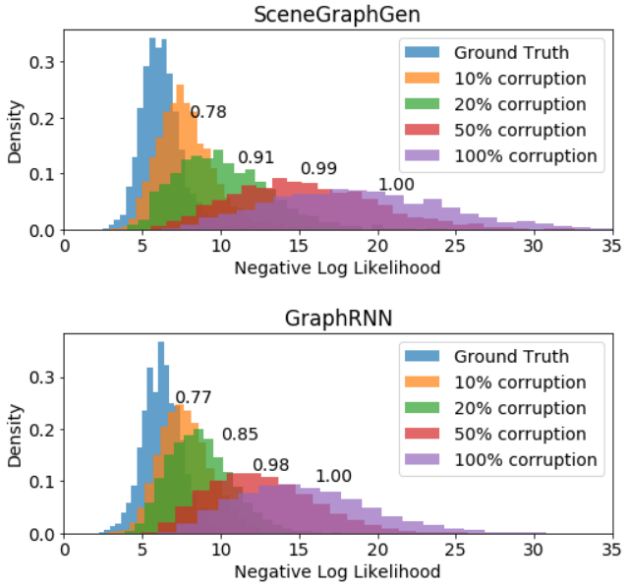


Figure 6. Distribution of NLL under varied levels of corruption

A.6. Checking for overfitting

Figure 7 shows examples of generated graphs and the respective nearest graph (via graph kernel comparison) from training data. The graphs are not identical, *i.e.* the model is not reproducing examples from the training set.

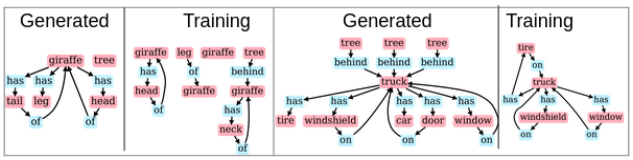


Figure 7. Examples of generated graphs as well as closest sample from the training set.

References

[1] S. Boughorbel, Jean-Philippe Tarel, and Francois Fleuret. Non-mercer kernels for svm object recognition. 01 2004.

[2] Matthew Fisher, Manolis Savva, and Pat Hanrahan. Characterizing structural relationships in scenes using graph kernels. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, New York, NY, USA, 2011. Association for Computing Machinery.

[3] M. Hein and O. Bousquet. Kernels, associated structures and generalizations. Technical Report 127, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, July 2004.

[4] Md Rahman and Nizar Bouguila. Efficient feature mapping in classifying proportional data. *IEEE Access*, PP:1–1, 12 2020.