

1. はじめに

シーングラフを用いた画像生成では、オブジェクト間の関係をエッジとして表現したシーングラフを入力し、画像を生成する。代表的な手法である Image Generation from Scene Graph (sg2im) [1] は、シーングラフから各オブジェクトのレイアウトを作成し、画像を生成する。しかしながら、オブジェクトの色や向きなどの詳細な情報が不足すると画像を正しく生成できない。この問題は学習データに起因すると考えられる。そこで本研究では、学習時に使用する代表的なデータセットである visual genome (VG) [2] から得られるシーングラフを調査し、どのようなシーングラフの場合に正しく画像生成が可能であるかを明らかにする。

2. Image Generation from Scene Graph

sg2im [1] は、シーングラフから画像生成する代表的な手法である。図1に示すように、まずシーングラフの各ノードでグラフ畳み込み処理をすることで、各ノードの特徴量を得る。その後、特徴量から Bounding box と Segmentation mask を含んだシーンレイアウトを作成する。作成したシーンレイアウトを生成モデルに入力して画像を生成する。

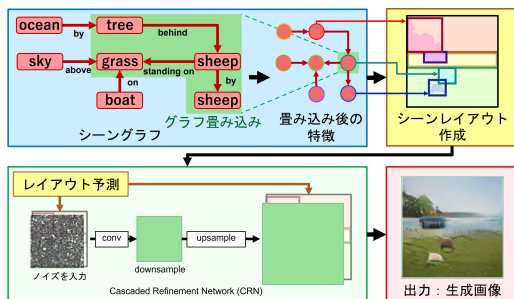


図 1 : sg2im の概略図

3. visual genome

VG [2] は、風景画像や動物など様々な画像を含むデータセットである。各画像に対して、オブジェクトに対する Bounding box と、オブジェクトの関係をシーングラフで表現したデータが含まれている。データ内には、オブジェクトを示す単語は 33,877 種類、関係を示す単語は 42,374 種類存在する。

4. 画像生成実験

本実験では、調査内容と画像生成実験に用いる VG データの調整、実験結果について説明する。

4.1. 調査内容

正しく画像を生成できないシーングラフの調査するために、以下の A. と B. の 2 項目について調査を行う。

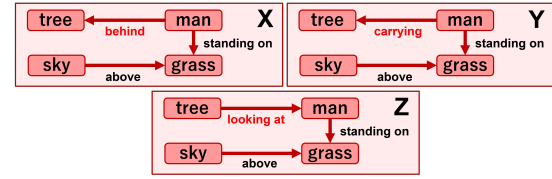
- 学習に用いるデータに含まれる総量やオブジェクト、関係の出現数の調査 (4.2 節)
- sg2im の事前学習済みモデルを用いて画像生成を行い、正しい画像を生成できるか確認 (4.3 節)

調査 B. で対象とするシーングラフを図 2 に示す。図 2 のシーングラフを用いて、どのようなシーングラフの場合に正しく画像生成が可能かを調査するために、以下の C. と D. の 2 項目の調査を行う。

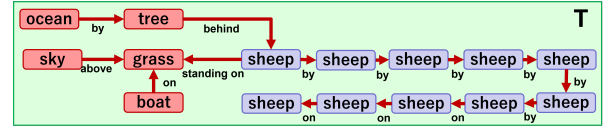
- データセット内の出現数が異なる関係に対して正しく画像が生成できるか (シーングラフ X, Y, Z)
- シーングラフ内のオブジェクト数を変えた時に正しく画像を生成できるか (シーングラフ T)

4.2. 学習データの調査結果 (A, B)

sg2im は、データを train, val, test に分割し、出現回数が極端に少ないオブジェクトと関係を削除する前処理を行う。残ったオブジェクトと関係の種類は 178 と 45 となり、1 つのサンプルにつきオブジェクトは平均 10 個、関係は平均 5 個である。処理後のデータの出現数を図 3 に示す。



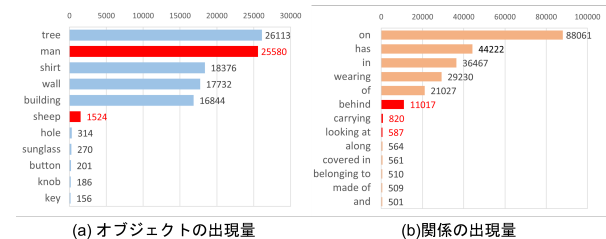
(a) 出現数の異なる関係の影響を評価するシーングラフ



(b) オブジェクト数の影響を評価するシーングラフ

図 2 : 実験に用いるシーングラフ

図 3 では、各オブジェクト、関係の出現数を上位 5 個と下位 5 個を示している。また、今回の実験で着目するものを赤色で示している。



(a) オブジェクトの出現量

(b) 関係の出現量

図 3 : オブジェクト、関係の合計出現数

4.3. 画像生成の調査 (C, D)

図 2(a)X, Y, Z のシーングラフによる生成結果を図 4X, Y, Z に示す。図 4X, Y, Z の生成結果から、オブジェクト間の関係を変えたシーングラフを入力しても、大まかに理解可能な画像が生成されている。よって、データセット内の出現数は生成品質に影響しないと考えられる。次に、図 2(b)T のシーングラフによる生成結果を図 4T に示す。図 4T の生成結果から、1 つのシーングラフに含まれるオブジェクト数が増加すると背景に存在する木が正しく生成できていない。また、どの生成でも羊の向きは変化せず単調な生成であることを確認した。よって、1 つのシーングラフに使用するオブジェクト数が生成画像の品質に影響することが分かった。これはオブジェクトが増えることで関係も増え、シーングラフが複雑になるためであると考えられる。

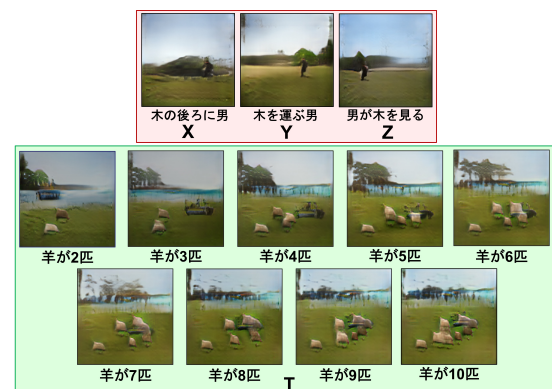


図 4 : 図 2 のシーングラフによる生成結果

5. おわりに

本研究では、sg2im による学習データ処理後の内容調査を行った。実験より、1 つのシーングラフのオブジェクト数が生成画像に影響することを確認した。今後は、他データセットの調査や今回発見した問題点の改善方法を模索する。

参考文献

- [1] J. Johnson *et al.*, "Image generation from scene graphs." CVPR, 2018.
- [2] R. Krishna *et al.*, "Visual genome: Connecting language and vision using crowd-sourced dense image annotations." IJCV, 2017.