

意味構造に着目したシーングラフ生成手法の提案

Scene Graph Generation Focusing on Semantic Structure

学籍番号：201821601

氏名：嵐 一樹

Arashi Kazuki

画像内容を構造的に理解することは、コンピュータビジョンの分野での重要な課題である。画像内容を構造的に理解するとは、画像中に示されている物体同士の関係や物体の状態などを、人間が視覚的に判断する結果と同等に理解するということである。

画像を理解する手法のひとつとしてシーングラフがある。シーングラフは画像内容を構造的に表し、グラフ構造を用いることで物体同士の関係性を明らかにすることに長けている。シーングラフでは、画像内容を物体情報、物体に付与されている属性情報、物体ペア間の関係性情報の3要素によって捉える。本研究ではわかりやすさのため、物体情報を Objects、物体に付与されている属性情報を Attributes、物体ペア間の関係性情報を Relationships と表記する。シーングラフは Objects をノードとし、Relationships をエッジとしたグラフ構造をしている。Attributes は Objects に付与する形の終端ノードとなっている。

シーングラフに関する研究は多々あり、シーングラフ自体の生成が主要な分野である。今までは、画像とシーングラフのペアデータセットを元にしてシーングラフを生成する研究が主であった。しかし、現在のシーングラフ生成は画像に強く依存しているため、画像から認識した情報を出来るだけ多く取り出した膨大なグラフ構造となっている。そこで、シーングラフとキャプション情報を用いることで画像への依存を減らし、シーングラフの改良ができると推測される。画像への依存を減らすことで、より画像の特徴的な情報を捉えることを目指した。

シーングラフの改良を行なうために、いくつかの手法を提案する。まずは、画像の本質的な情報を抽出する手法である。シーングラフにおけるノードとエッジの出現確率を考えることで、グラフ上の本質的な情報を捉えることができると考えられる。画像の特徴を明確に表している要素を抽出することで、より精度の高いシーングラフの生成を目指す。次に、シーングラフを簡潔な表現にすることで、これは表現を簡潔にすることで、よりシーングラフの意味が伝わりやすくと考えたからである。同じ要素が複数回出現している部分グラフに対して、不要なノードとエッジを削除することを目指す。これらの提案手法を行なった結果、シーングラフがどのように変化するかを観測する。

提案手法を評価するための実験を行なう。評価実験のデータセットとしては、Visual Genome と MSCOCO(Microsoft Common Objects in Context) を用いる。Visual Genome は画像とシーングラフがペアとなったデータセットであり、画像毎に Objects, Attributes, Relationships の情報や小領域ごとのキャプションが含まれている。MSCOCO は、画像とそれに対応するキャプションが付与されているデータセットである。これらを組み合わせたデータセットを作成し、評価実験に使用する。また、シーングラフは単純な有向グラフではないため、改良を評価するための手法が存在しない。そのため、シーングラフの改良を示すための手法を新たに導入し、評価実験の結果と共に示す。

評価実験の結果、より簡潔なシーングラフが生成できることを示し、より画像内容を表すシーングラフを作成することができた。

研究指導教員：三河 正彦

副研究指導教員：若林 啓