

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	知能コミュニケーション (中村 哲 (教授))		
学籍番号	2211241	提出日	令和 6年 1月 19日
学生氏名	濱田 裕太		
論文題目	シーングラフを用いたText-to-Imageの評価		
要旨			
<p>Vision&Language分野において指示文に従った画像を生成するText-to-Imageは、言語と画像の融合を目的としたマルチモーダル学習のタスクの一つである。既存モデルの課題として指示文で指定された物体間の関係や数量といった空間的構造を生成画像に正確に反映することを苦手とする点があり、これを解決するためには空間的構造の把握に適した画像生成評価指標が必要である。そこで本研究では、シーングラフを用いたText-to-Imageの評価手法を提案することで問題の解決を試みる。具体的には、指示文と生成画像からそれぞれのシーンを表すシーングラフを生成し、生成された二つのグラフを比較することにより空間的構造の一貫性評価を行う。</p> <p>そして提案手法の有効性を検証するため、実画像と2種の生成画像、指示文を模したテキストで構成されるデータセットを作成し、実験を行った。実験の結果、生成画像を対象としても物体検出は可能であること、しかし既存のSGG・PSG手法ではバイアス問題により関係予測の精度が低く、トリプレット(主語-関係語-目的語)の比較を利用する本手法を空間的構造の評価に利用するには関係予測モジュールの改善が必要ながわかった。さらに、提案した2手法の結果の差から、画像からのシーングラフ生成には、一般的なSGGを用いた手法に比べて指示文中の物体に注目したPSGによる手法の方が適していることがわかった。</p>			