

幾何学図形を用いた合成データによる 大規模視覚言語モデルの視覚推論能力向上

大阪大学 情報科学研究科 大倉 史生・對野 裕己

目的 大規模視覚言語モデル（LVLM）において、視覚情報を十分に活用できず言語情報に過度に依存する問題を解決し、視覚中心タスクの性能向上を目指す。

内容 複雑な幾何学図形からなる合成データセットを用い、実画像と混合して学習することで、LLMが視覚情報をより深く利用できるようにする手法を提案する。

結果 提案手法により視覚中心ベンチマークで精度が向上し、特にLLM側の学習が有効であることを確認した。

利用した計算機 SQUID GPUノード群
ノード時間 14,000 時間
使用ソフトウェア PyTorch
1ノード8GPU並列

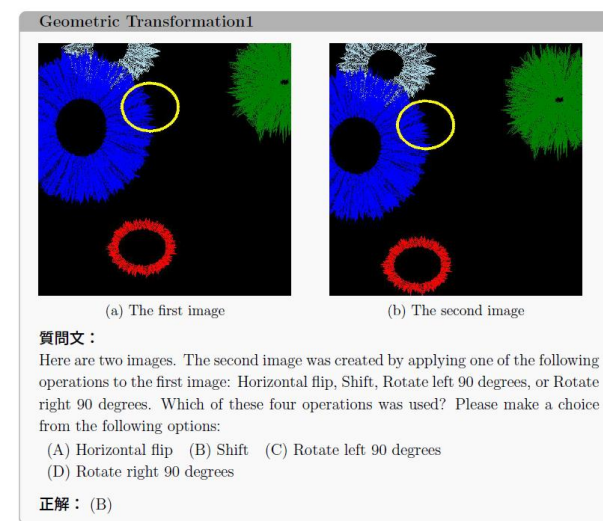


図: 合成データによる
学習タスク例