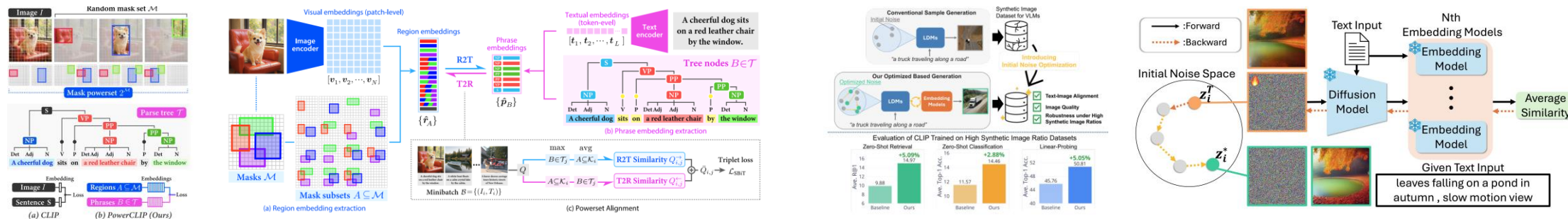


課題名：
データの倫理問題を根本解決する視覚言語基盤モデル構築

実施時期：2025年4月-2026年3月
所属機関名：産業技術総合研究所
代表者氏名：原 健翔

成果概要：
視覚言語基盤モデルの構築では、Web由来の大規模画像・テキストデータに依存するため、著作権、個人情報、バイアス、データ由来の不透明性が大きな課題となる。本研究では、実画像データへの依存を低減しつつ高性能な視覚言語モデルを構築するため、合成データ生成、細粒度な画像言語対応学習、およびデータ分布評価・選別の3つの基盤技術を開発した。

成果のポイント：



本課題では、視覚言語基盤モデルの学習における大規模Webデータ依存を低減するため、合成データ生成、データ評価、細粒度な事前学習を統合した新しいモデル構築基盤を開発した。従来は、実世界から収集した大量の画像・テキストペアに依存していたため、著作権、個人情報、バイアス、データ由来の不透明性が避けがたい課題であった。

これに対し、本研究では、拡散モデルにより生成した合成画像をVLM学習に有効活用する手法、データの有用性や分布差をCLIP型モデルにより評価する手法、さらに画像領域と言語句の細粒度対応を学習するPowerCLIPを開発した。これらにより、単に大規模データを収集するのではなく、**安全に生成し、数理的に選別し、高精度に学習する**視覚言語基盤モデル構築の枠組みを実現した。

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：

https://openaccess.thecvf.com/content/WACV2026/html/Ohkubo_Diffusion_Noise_Optimization_for_Synthetic_VLM_Training_WACV_2026_paper.html

<https://arxiv.org/abs/2511.23170>

<https://arxiv.org/abs/2506.22881v2>