

課題名：データの枯渇問題を軽減する3D基盤モデル構築

実施時期：2025年4月～2026年3月

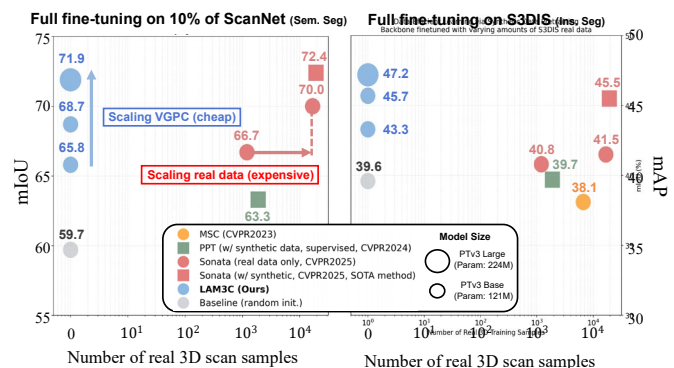
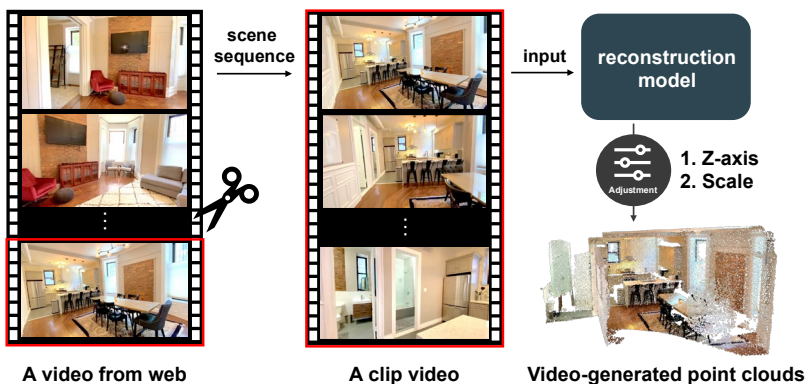
所属機関名：産業技術総合研究所

代表者氏名：原 健翔

成果概要：3D空間を高精度に認識・理解するAIモデルの構築には、大量の3Dデータが必要である。しかし、3Dデータの収集には専用機器や手作業が必要となる場合が多く、収集コストが高い。そのため、既存の3Dデータセットは、テキストや画像のデータセットと比べて規模が限られている。そこで本研究では、実環境で3D点群を直接スキャンすることなく、Web上に存在する動画を活用して3D認識モデルを構築する手法を提案した。具体的には、YouTube上の屋内シーン動画を収集し、3D再構成モデルを用いて屋内3Dシーンデータを生成する。さらに、生成したデータを自己教師あり学習によって学習することで、3Dシーン理解に有効なAIモデルを構築した。実験の結果、屋内シーンの領域分割タスクにおいて従来手法を上回る認識精度を達成し、Web動画が3D認識モデルの学習データとして有効であることを示した。

成果のポイント：教師なし動画から実空間を認識する3D視覚表現学習を実現

- RoomTours Dataset (屋内シーン擬似点群データセット)
 - Youtubeから屋内環境動画を収集し、3D再構成モデル(VGGT)を用いて擬似点群を自動生成
- LAM3C (擬似点群に対する自己教師あり学習)
 - 擬似点群シーンに含まれるノイズ点群などに対する表現学習の安定性を改善する損失関数を提案



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：

Ryosuke Yamada, Kohsuke Ide, Yoshihiro Fukuhara, Hirokatsu Kataoka, Gilles Puy, Andrei Bursuc, Yuki M. Asano, “3D sans 3D Scans: Scalable Pre-training from Video-Generated Point Clouds,” CVPR, 2026.