

ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名：動画像における柔軟な転移学習を実現する事前学習モデルの構築	実施時期：2025/1/20 ~ 2025/3/31 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：本研究ではアノテーションコストが静止画以上に大きい動画像データにおける柔軟な転移学習を可能とするために、大規模かつ多様なデータベースを利用して事前学習モデルを構築した。2つの数式ドリブンデータベースと3つの公開実動画データベースを統合し組み合わせて利用する事前学習をABCI3.0にて実施し学習済みモデルを構築した。また、構築したモデルの性能評価を5つのドメイン・18のデータセットからなるベンチマークにより実施した。	
成果のポイント： 本研究では動画像における柔軟な転移学習を可能とすることを目的として、事前学習モデルの構築および性能評価を実施しそれらを公開するという成果を挙げた。 事前学習モデルの構築においては、まず2つの数式ドリブンデータベースをそれぞれ利用してモデルの初期学習を実施した。具体的には、2D画像と3D点群のペアデータを人工的に生成することで構築されたVG-FractalDBにより空間的な表現を、動画像データを人工的に生成することで構築されたVideo Perlin Noiseデータセットにより時間的な表現を学習させた。加えて、初期学習後のモデルに対して、Kinetics-700, Moments in Time, Something-Something v2という3つの公開実動画データベースを統合して利用する事前学習を実施した。このような多様なドメイン、多様なカテゴリを含む実動画データベースでの学習により事前学習モデルの汎用性を向上させることを図った。事前学習モデルとしては、畳み込みニューラルネットワーク（CNN）ベースのモデル1種と、Transformerベースのモデル2種を用意し、利用する計算リソースに応じて適切なモデルを選択可能とした。 また、5つのドメイン・18のデータセットから構成されるBEARベンチマークを用いて各モデルの性能を評価し、利用者が目的とするドメインにおける性能を把握できるようにした。これらの事前学習モデルと性能評価結果はまとめて公開し、計算環境に応じて選択可能な複数サイズのモデルを多様なドメインでの評価結果とともに提供することで、社会実装の容易化を促進することを狙った。	
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： モデル公開ページ： https://github.com/kenshohara/video-recognition-models	