

2023年 2月 14日

豊橋技術科学大学長 殿

情報・知能工学専攻
学位審査委員会
委員長 岡田 美智男



論文審査及び最終試験の結果報告

このことについて、博士学位論文審査を実施し、下記の結果を得ましたので報告いたします。

学位申請者	Yuri Yudhaswana Joeфриe		学籍番号	第 179304 号
申請学位	博士 (工学)	専攻名	大学院工学研究科博士後期課程 情報・知能工学 専攻	
博士学位論文名	On Spatio-Temporal and Motion Feature Representation Learning with Deep Neural Network for Accurate Video Action Recognition (正確なビデオアクション認識のための深層ニューラルネットワークを用いた時空間と動作特徴量の表現学習について)			
論文審査の期間	2023年1月12日 ~ 2023年2月14日			
公開審査会の日	2023年2月14日	最終試験の実施日	2023年2月14日	
論文審査の結果※	合格		最終試験の結果※	合格
審査委員会(学位規程第6条)				
学位申請者にかかる博士学位論文について、論文審査、公開審査会及び最終試験を行い、別紙論文内容の要旨及び審査結果の要旨のとおり確認したので、学位審査委員会に報告します。				
委員長	北岡 教英			
委員	栗山 繁		青野 雅樹	
		印		印
		印		印

※論文審査の結果及び最終試験の結果は「合格」又は「不合格」の評語で記入すること。

論文内容の要旨

SNS上で動画コンテンツが急激に増大している。AIによる動画の自動解析技術のニーズも急激に増大しているが、個々の動作の種類を、軽量かつ細粒度で認識する技術は、まだ開発途上である。本論文では、動画データセットに対して、新たなニューラルネットを構築し、学習時間が少なく、軽量だが、動作の細粒度での高精度な認識が可能な提案手法を開発し、動作認識でよく用いられる複数のデータセットで、その有効性・優位性を示している。

第1章では、動画分類の背景・動機を、第2章では、動画分類の関連技術、それに関わる深層学習技術を述べている。第3章では、提案手法の開発動機となった特徴学習 (feature learning) に関してサーベイしている。

第4章と第5章で提案手法を述べている。第4章では、日常生活の動作データを主とする3つの動画データセットを用いて、時空間データ表現から、空間方向、空間縦方向と時間方向、空間横方向と時間方向の3つに分割するマルチビュー手法という手法を含む提案手法を述べている。同時に、従来のオプティカルフローに代表される低速で計算量の多い時間変化を捉える手法に代わり、連続フレームの差分を、直接ニューラルネットアーキテクチャ内で処理する仕組みを中心とした手法を述べている。第5章では、時間的混合ガウス法 (TGM: Temporal Gaussian Mixture) 技術に着目し、これを構成要素として、時空間データをフュージョンする3種類の新しいアーキテクチャを導入し、スポーツ動画に多い、同時に複数の動作が発生する場合のマルチラベル・マルチクラス問題に焦点をあて、従来手法よりも高精度を達成したことを報告している。第6章では、まとめを述べるとともに、今後の課題について述べている。

審査結果の要旨

動画コンテンツは、YouTube、Netflixのような動画に特化したSNSや配信サービスに限らず、オリンピック、様々なスポーツの世界大会など個別な動画配信サービス、個人のTikTok、Instagram、Twitterなどでも非常に多く普及している。静止画と比べて動画は、一般にデータ量が多く、コンテンツとして描写される対象物 (オブジェクト) の動きも、シンプルなものから、同時に複数の動作が発生する複雑なものまで多種多様である。

本論文では、動画中での動作認識を2種類の分類問題として捉えている。ひとつは、マルチクラス分類問題、もうひとつはマルチラベル・マルチクラス問題である。マルチラベル問題は、同時に複数の動作が発生するため、難度の高い問題である。前半のマルチクラス問題に対して、動画を、 T (時間軸) $\times H$ (空間縦軸) $\times W$ (空間横軸) のデータとして表現し、これを $T \times H$ 、 $T \times W$ 、 $H \times W$ の3つに分解することで、軽量性を達成し、同時に正確な動作認識を行うユニークな手法を提案した。様々な日常動作が含まれる3種類の動画のベンチマークデータセットにおいて、従来手法との比較実験でその有効性、優位性を示している。

一方、マルチラベル・マルチクラス問題に関しては、細粒度の高精度な分類精度を達成できる技術を参照してTGM (Temporal Gaussian Mixture) を構成要素とする3種類の深層学習アーキテクチャを提案し、同時複数動作の多いスポーツ動画のデータセットで、従来手法よりも高精度な分類性能を達成している。具体例として、学術的によく使用されるベンチマークデータセットに含まれるバレーボールのjumping、standing、spiking、runningなど同時性が高く、見た目も似た動作識別を、マルチラベル・マルチクラス問題として実証し、高精度な識別能力を提示している。

これらの成果は、瞬間的な動作を見極められる (スポーツ実況などの現場で) 専門家の介入を必要としないで、軽量かつ高性能な動作の自動認識手法を提案している点で特筆すべきものであり、学術的に高い評価を受けるとともに、動画認識分野での幅広い応用が期待される。以上より、本論文は博士 (工学) の学位論文に相当するものと判定した。