

論文審査の要旨および学識確認結果

報告番号	甲 第 号	氏 名	鈴木 智之
論文審査担当者：	主査	慶應義塾大学 教授	博士（工学） 青木 義満
	副査	慶應義塾大学 教授	工学博士 池原 雅章
		慶應義塾大学 教授	博士（工学） 杉本 麻樹
		慶應義塾大学 専任講師	博士（工学） 吉岡 健太郎
学士（工学）、修士（工学）鈴木智之君提出の学位請求論文は「Efficient Video Recognition with Transformers: Tackling the Redundancy in Video」（Transformer を用いた効率的動画認識：動画の冗長性への取り組み）と題し、6章から構成されている。 動画は実世界の連続的な時間の流れの中で観測される視覚データである。コンピュータビジョン分野において、動画から解釈可能な情報を抽出する動画認識は、行動解析、危険予測、動画検索など、幅広い応用をもつ基盤技術として注目されている。一方で、その計算コストの大きさが実応用上の障害となっているのが現状である。我々は映写機の時代から、動画を画像の系列として表現・可視化してきた。そして現代の動画認識技術は、同じようにコンピュータに動画を画像の系列として順次処理させるのが主流である。しかし、動画の時間的連続性を考慮すると、動画の微小時間区間における変化は非常に小さく、これをすべて同等に処理することは非常に冗長であると考えられる。 本論文は、動画の冗長性に着目し、それを削減することで、動画認識の効率化を目指している。近年、動画認識において高い性能を示している Transformer をベースとするモデルを想定し、それらの内部で処理される冗長なトークンを検出・除外することで、動画認識を効率化する手法を提案している。すべての時刻が観測可能なオフライン動画認識と、動画が刻一刻と観測されるオンライン動画認識の二つのシナリオに適した手法をそれぞれ提案している。いずれの設定においても、広範な実証実験を行い、提案手法によって精度を維持しながら、動画認識の処理速度を大幅に向上可能であることを示している。 第1章では、動画や動画認識、動画の冗長性に関する問題を提起し、本論文の研究課題と貢献を示している。 第2章では、動画認識の歴史とその効率化の試みについて概説し、その歴史的な流れの中で本研究の位置付けを明らかにしている。 第3章では、本研究の基盤となる Transformer とそれを用いた動画特徴抽出についての予備知識を説明している。 第4章では、オフライン動画認識において、動画の圧縮情報を活用して冗長トークンを検出・除外する手法を提案している。オフライン動画認識のベンチマークである行動認識において評価を行い、その効果を実験的に示している。 第5章では、圧縮情報が利用できないオンライン動画認識において、過去の情報を参照して現時刻フレームの冗長トークンを検出し、過去情報を再利用することでそれらの情報処理をスキップする手法を提案している。さらに削減対照を入力トークンだけでなく、中間トークンに拡張する手法を提案している。オンライン動画認識における複数のタスクにおいて提案手法の評価を行い、その有効性を実証している。 第6章では、本論文を総括し、今後の展望について議論している。 以上要するに、本論文は、コンピュータビジョンにおける動画認識タスクを対象に、動画の冗長性に着目することで、Transformer を用いた動画認識モデルの効率化を実現したものであり、画像工学分野において、工学上、工業上寄与するところが少なくない。よって、本論文の著者は博士（工学）の学位を受ける資格があるものと認める。			
学識確認結果	学位請求論文を中心にして関連学術について上記審査会委員および総合デザイン工学特別研究第2（電気電子工学専修）科目担当者で試問を行い、当該学術に関し広く深い学識を有することを確認した。 また、語学（英語）についても十分な学力を有することを確認した。		