論　文　内　容　の　要　旨

博士論文題目　　Actor-identified Spatiotemporal Action Detection—
Detecting Who Is Doing What in Videos
（行動中の人物を同定した時空間行動検出---
動画中の誰が何をしているかの検出）

氏　　名　　　　Fan Yang

（論文内容の要旨）

The success of deep learning on video Action Recognition (AR) has motivated researchers to progressively promote related tasks from the coarse level to the fine-grained level. Compared with conventional AR that only predicts an action label for the entire video, Temporal Action Detection (TAD) has been investigated for estimating the start and end time for each action in videos. Taking TAD a step further, Spatiotemporal Action Detection (SAD) has been studied for localizing the action both spatially and temporally in videos. However, who performs the action, is generally ignored in SAD, while identifying the actor could also be important. To this end, we propose a novel task, Actor-identified Spatiotemporal Action Detection (ASAD), to bridge the gap between SAD and actor identification.

In ASAD, we not only detect the spatiotemporal boundary for instance-level action but also assign the unique ID to each actor. To approach ASAD, Multiple Object Tracking (MOT) and Action Classification (AC) are two fundamental elements. By using MOT, the spatiotemporal boundary of each actor is obtained and assigned to a unique actor identity. By using AC, the action class is estimated within the corresponding spatiotemporal boundary. Since ASAD is a new task, it poses many new challenges that cannot be addressed by existing methods: i) no dataset is specifically created for ASAD, ii) no evaluation metrics are designed for ASAD, iii) current MOT performance is the bottleneck to obtain satisfactory ASAD results. To address those problems, we contribute to i) annotate a new ASAD dataset, ii) propose ASAD evaluation metrics by

considering multi-label actions and actor identification, iii) improve the data association strategies in MOT to boost the MOT performance, which leads to better ASAD results. We believe considering actor identification with spatiotemporal action detection could promote the research on video understanding and beyond.

氏　名	Fan Yang

（論文審査結果の要旨）

To detect who is doing what in videos is one of the important technologies for video-analysis, human-interaction, and surveillance. The current state-of-the-art technologies made remarkable progress on video action recognition (AR) by the success of deep learning. AR only predicts an action label for the entire video. Temporal Action Detection (TAD) estimates the start and end time for each action in videos. Spatiotemporal Action Detection (SAD) localizes the action both spatially and temporally in videos. In this research, he developed datasets and methods for Actor-identified Spatiotemporal Action Detection (ASAD) integrating AR, TAD, SAD, multiple object tracking (MOT), and action classification (AC) into one system.

To address those problems, he contributed to i) annotate a new ASAD dataset, ii) propose ASAD evaluation metrics by considering multi-label actions and actor identification, iii) improve the data association strategies in MOT to boost the MOT performance, which leads to better ASAD results. In particular, the proposed method not only detects the spatiotemporal boundary for instance-level action but also assigns the unique ID to each actor. To realize ASAD, Multiple Object Tracking (MOT) and Action Classification (AC) are integrated into the ASAD system. By using MOT, the spatiotemporal boundary of each actor is obtained and assigned to a unique actor identity.

Actor-identified Spatiotemporal Action Detection is challenging tasks and inevitable technologies in video-analysis, human-interaction, and surveillance. This thesis research proposed solutions to these problems, and a series of his research resulted in four high quality peer-reviewed international journal papers, and three peer-reviewed international conference papers. Further, he was awarded by the top-ranking scores in two international challenges. As a result, the thesis is sufficiently qualified as a Doctoral thesis of Engineering.