

# PHÂN ĐOẠN THỰC THỂ TRÊN VIDEO SỬ DỤNG TƯ THỂ CỦA NGƯỜI VÀ TĂNG CƯỜNG DỮ LIỆU

*Trần Minh Triết<sup>1</sup>, Ninh Văn Tú<sup>1</sup>, Lê Tử Khiêm<sup>1</sup>, Tôn Thất Vĩnh<sup>1</sup>,  
Nguyễn Khắc Tuấn<sup>1</sup>, Lê Trung Nghĩa<sup>2</sup>, Nguyễn Văn Tâm<sup>3</sup>*

<sup>1</sup>Khoa Công nghệ Thông tin,  
Trường Đại học Khoa học Tự Nhiên, ĐHQG-HCM

<sup>2</sup>Viện Tin học quốc gia Nhật Bản

<sup>3</sup>Đại học Dayton, Hoa Kỳ

[tmtriet@fit.hcmus.edu.vn](mailto:tmtriet@fit.hcmus.edu.vn), [nvtu@apcs.vn](mailto:nvtu@apcs.vn), [ltkhiem@apcs.vn](mailto:ltkhiem@apcs.vn), [ttvinh@selab.hcmus.edu.vn](mailto:ttvinh@selab.hcmus.edu.vn),  
[nktuan@apcs.vn](mailto:nktuan@apcs.vn), [ltnghia@nii.ac.jp](mailto:ltnghia@nii.ac.jp), [tamnguyen@udayton.edu](mailto:tamnguyen@udayton.edu)

## Tóm tắt

Phân đoạn thực thể trong video là một trong những bài toán thú vị nhưng còn nhiều vấn đề thách thức cần giải quyết. Trong bài báo này, nhóm tác giả đề xuất phương pháp mới cho việc phân đoạn thực thể trong video với hai thành phần chính. Trước tiên, chúng tôi đề xuất giải pháp để sử dụng thông tin tư thể của người để định hướng việc phân đoạn thực thể người. Ngoài ra, chúng tôi tổng hợp dữ liệu mới và huấn luyện các mô hình FCN, bao gồm cả DeepLab2 và OSVOS, để tách các thực thể không phải là người. Phương pháp được đề xuất đạt kết quả cao trên tập dữ liệu DAVIS 2017 và là một trong những phương pháp đạt kết quả cao nhất trong cuộc thi phân đoạn thực thể video 2018.

Từ khóa: phân đoạn thực thể trong video, sử dụng tư thể người, tăng cường dữ liệu.

# INSTANCE SEGMENTATION IN VIDEO WITH HUMAN-POSE GUIDANCE AND DATA AUGMENTATION

*Minh-Triet Tran<sup>1</sup>, Tu V. Ninh<sup>1</sup>, Tu-Khiem Le<sup>1</sup>, Vinh Ton-That<sup>1</sup>,  
Khac-Tuan Nguyen<sup>1</sup>, Trung-Nghia Le<sup>2</sup>, Tam Nguyen<sup>3</sup>*

<sup>1</sup>Faculty of Information Technology, University of Science, VNU-HCM

<sup>2</sup>National Institute of Informatics, Japan

<sup>3</sup>University of Dayton, U.S.A.

[tmtriet@fit.hcmus.edu.vn](mailto:tmtriet@fit.hcmus.edu.vn), [nvtu@apcs.vn](mailto:nvtu@apcs.vn), [ltkhiem@apcs.vn](mailto:ltkhiem@apcs.vn), [ttvinh@selab.hcmus.edu.vn](mailto:ttvinh@selab.hcmus.edu.vn),  
[nktuan@apcs.vn](mailto:nktuan@apcs.vn), [ltnghia@nii.ac.jp](mailto:ltnghia@nii.ac.jp), [tamnguyen@udayton.edu](mailto:tamnguyen@udayton.edu)

## **Abstract**

Instance segmentation in video is a new interesting yet challenging problem. In this work, we introduce a novel method for instance segmentation with two main components. First, we propose to use human pose as the guidance for human instance segmentation. Second, we synthesize data to train fully convolutional networks to extract non-human objects. Our hybrid method achieves high results on DAVIS 2017 dataset, among the state-of-the-art methods of the DAVIS challenge on Video Object Segmentation 2018..

Key words: video instance segmentation, human pose guidance, data augmentation