UNIVERSITY OF OTTAWA

DOCTORAL THESIS

Thesis Title

Author:
Nicolas FLEECE

Supervisor:
Dr. Robert LAGANIÈRE

A thesis submitted in fulfillment of the requirements
for the degree of Masters of Computer Science, Specialization in Applied Artificial Intelligence

in the

VIVA Lab
Department of Computer Science

July 23, 2022

UNIVERSITY OF OTTAWA

Abstract

Engineering

Department of Computer Science

Masters of Computer Science, Specialization in Applied Artificial Intelligence

Thesis Title

by Nicolas FLEECE

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

QUOTE HERE

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor…

Contents

List of Figures

List of Tables

List of Abbreviations

LAH **L**ist **A**bbreviations **H**ere

WSF **W**hat (it) **S**tands **F**or

Chapter 1

Literature Review

1.1 Classical Action Recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

1.2 Deep Learning Action Recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

1.2.1 3D-CNN

ResNet

MoveNet

1.3 Optical Flow

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

1.4 Person-Based Action Recognition

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

1.5 Pose-Based Action Recognition

Pose involves extracting the skeleton of the person and using this data over multiple frames of a video to classify an action. Pose is a common addition used in action recognition as it relates most to how humans view actions and the movement of different bones.

1.5.1 Intermediate Representations

The approach of the majority of this thesis involves creating intermediate representations for pose data over multiple frames. This typically has the aim of creating some kind of image that represents either the motion of the persons bones and/or joints through the image at different points in the video. These images can then be used either by the model independently or added to traditional two-stream architectures.

PoTion

Pose motion representation for action recognition [1] was largely the inspiration for most of the work that was done within the thesis. This approach aims to take the joints extracted from the pose representation and use the movement over f frames, creating j images where j is the number of joints.
The construction of the intermediate representations is based off of joint probability locations that are provided through pose estimation models.

Bibliography

[1] V. Choutas, P. Weinzaepfel, J. Revaud, and C. Schmid, "Potion: Pose motion representation for action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7024–7033.