La borsa di dottorato è stata cofinanziata con risorse del
Programma Operativo Nazionale Ricerca e Innovazione 2014-2020 (CCI 2014IT16M2OP005),
Fondo Sociale Europeo, Azione I.1 "Dottorati Innovativi con caratterizzazione Industriale"

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI SCIENZE AZIENDALI – MANAGEMENT E INNOVATION SYSTEMS

Dottorato di Ricerca in Big Data Management

XXXIII ciclo

Multiple Object Tracking and Face-based Video Retrieval: Applications of Deep Learning to Video Analysis

Relatore:

Ch.mo Prof.

Roberto TAGLIAFERRI

Candidato:

Gioele

CIAPARRONE

Mat. 8801500009

Coordinatore:

Ch.mo Prof.

Valerio ANTONELLI

ANNO ACCADEMICO 2019/20

Abstract

In recent years, deep learning (DL) has obtained numerous successes in analyzing complex data, such as images or audio. A particularly recent area of application is the analysis of videos.

This thesis focuses on the application of deep learning algorithm to two video analysis tasks: Multiple Object Tracking (MOT) and Face-based Video Retrieval (FBVR).

The first main part of the thesis presents an in-depth survey of the state of the art of DL-based MOT algorithms. This is the first comprehensive survey specifically on the use of DL for MOT, focusing on 2D frames extracted from single-camera videos. I identify the four main steps of a MOT algorithm and describe the various DL techniques used in the literature in each of those four steps. I also collect and compare results obtained by existing algorithms on the most common MOT datasets and I analyze the most successful techniques employed. Finally, I present a discussion about the open issues of current MOT algorithms and the possible solutions and future directions of research.

The second part of the thesis focuses instead on the task of FBVR. I present a novel pipeline for the retrieval of unconstrained multi-shot videos using faces, specifically in the context of television-like videos. Since no existing dataset in the literature is appropriate for an end-to-end evaluation of the proposed pipeline, I build a large-scale video dataset by adapting the VoxCeleb2 dataset to the task of FBVR. I compare and evaluate numerous DL-based approaches for the various steps in pipeline, such as shot detection, face detection and face recognition, and I describe the advantages and disadvantages of each employed technique. The best-performing configuration

of the pipeline obtains 97.25% Mean Average Precision on the independent test set, while performing each query on thousands of videos in less than 0.5 seconds. Finally, I describe the integration of the presented pipeline into the commercial software TVBridge, developed by CEDEO.