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

Negli ultimi anni il deep learning (DL) ha avuto molto successo nell'analisi di dati complessi, quali immagini o audio. Un'area di applicazione particolarmente recente è l'analisi di video.

Questa tesi tratta dell'applicazione di algoritmi di deep learning a due task di analisi video: il Multiple Object Tracking (MOT) e il Face-based Video Retrieval (FBVR).

La prima parte della tesi presenta un'approfondita revisione della letteratura sullo stato dell'arte di algoritmi MOT basati su DL. Questa è la prima revisione della letteratura a concentrarsi specificamente sull'utilizzo del DL per il MOT, in particolare per frame 2D estratti da video registrati con una singola videocamera. Ho identificato i quattro principali passi di un algoritmo MOT e descritto le varie tecniche di DL utilizzate in letteratura per ciascuno di questi quattro passi. Ho raccolto e confrontato i risultati ottenuti da algoritmi in letteratura sui più comuni dataset MOT e ho analizzato le migliori tecniche utilizzate. Presento infine una discussione riguardo ai problemi aperti degli algoritmi MOT esistenti, insieme alle possibili soluzioni e alle direzioni future di ricerca.

La seconda parte della tesi si concentra invece sul task del FBVR. Ho presentato una pipeline innovativa per la ricerca di video multi-shot senza restrizioni (*unconstrained*) tramite l'utilizzo di facce, nel contesto specifico di video di tipo televisivo. Poiché nessun dataset esistente in letteratura era appropriato per una valutazione esaustiva della pipeline proposta, ho costruito un dataset di video di grandi dimensioni riadattando il dataset VoxCeleb2 al task del FBVR. Ho confrontato e valutato diversi approcci basati su DL per

i vari passi della pipeline, tra cui identificazione degli shot, identificazione delle facce e riconoscimento facciale. Ho inoltre descritto vantaggi e svantaggi di ciascuna tecnica utilizzata. La migliore configurazione della pipeline ha ottenuto una Mean Average Precision pari al 97.25% sul test set indipendente, il tutto eseguendo ciascuna query su migliaia di video in meno di 0.5 secondi. Ho infine descritto il processo di integrazione della pipeline nel software commerciale TVBridge, sviluppato da CEDEO.