



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

ROZPOZNÁVÁNÍ OSOB A JEJICH ČINNOSTI VE VIDEU Z BEZPEČNOSTNÍCH KAMER

RECOGNIZING PEOPLE AND THEIR ACTIVITIES IN VIDEO FROM SECURITY CAMERAS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

JURAJ SAMUEL SALOŇ

VEDOUCÍ PRÁCE
SUPERVISOR

doc. RNDr. PAVEL SMRŽ, Ph.D.

BRNO 2021

Zadání bakalářské práce



23868

Student: **Saloň Juraj Samuel**

Program: Informační technologie

Název: **Rozpoznávání osob a jejich činnosti ve videu z bezpečnostních kamer**
Recognizing People and Their Activities in Video from Security Cameras

Kategorie: Umělá inteligence

Zadání:

1. Seznamte se s moderními metodami identifikace a sledování osob a klasifikace aktivit pomocí neuronových sítí.
2. Shromážděte dostupné datové sady pro průběžné testování systému.
3. Na základě získaných poznatků navrhněte a implementujte systém, který dokáže anotovat osoby a jejich chování ve videu z bezpečnostních kamer.
4. Vyhodnoťte výsledky systému na reprezentativním vzorku dat.
5. Vytvořte stručný plakát prezentující práci, její cíle a výsledky.

Literatura:

- dle domluvy s vedoucím

Pro udělení zápočtu za první semestr je požadováno:

- funkční prototyp řešení

Podrobné závazné pokyny pro vypracování práce viz <https://www.fit.vut.cz/study/theses/>

Vedoucí práce: **Smrž Pavel, doc. RNDr., Ph.D.**

Vedoucí ústavu: Černocký Jan, doc. Dr. Ing.

Datum zadání: 1. listopadu 2020

Datum odevzdání: 12. května 2021

Datum schválení: 7. dubna 2021

Abstrakt

Cieľom tejto práce je navrhnutie a implementovanie systému schopného rozpoznávať činnosti ľudí z bezpečnostných kamier. Hlavný dôraz je kladený na koncept komplexných situácií alebo udalostí, ktoré sú definované vzťahmi medzi rozpoznanými objektmi. Úvod tejto práce je venovaný oboznámeniu sa s jednotlivými časťami systému, kde sú objasnené využité techniky rozpoznávania objektov, sledovania objektov a rozpoznávania činností, ktoré sú použité v tejto práci. Druhá časť práce popisuje konečný návrh a implementáciu systému. Pre rozpoznávanie jednotlivých činností sú ďalej definované niektoré vzťahy medzi získanými informáciami popisujúcimi tieto činnosti, ako napríklad „vystupovanie z auta“ alebo „kráčanie dvoch a viacerých ľudí spolu“. Nakoniec je úspešnosť rozpoznávania vyhodnotená metrikou strednej priemernej presnosti (Mean Average Precision).

Abstract

The aim of this thesis is to design and develop a system capable of recognizing the activities of people from surveillance cameras. Special attention is paid to the concept of complex situations or events that are defined by relations between identified objects. The first part surveys state-of-the-art techniques for object recognition, object tracking, and recognition of activities relevant to the realized solution. The second part describes the design and implementation of the devised system. It takes advantage of specific relations among two or more objects that are identified in video recordings, such as “person getting out of the car” or “one or more people met with a person of interest and they left together”. Results are evaluated on video data extracted from available datasets and manually annotated. The mean average precision metric (MAP) on the data is reported.

Kľúčové slová

rozpoznávanie akcií, definícia činností, detekcia objektov, sledovanie objektov, analýza videa, neurónová siet

Keywords

action recognition, activity definition, object detection, object tracking, video analysis, neural network

Citácia

SALOŇ, Juraj Samuel. *Rozpoznávání osob a jejich činnosti ve videu z bezpečnostních kamer*. Brno, 2021. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce doc. RNDr. Pavel Smrž, Ph.D.

Rozpoznávání osob a jejich činnosti ve videu z bezpečnostních kamer

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením pána doc. RNDr. Pavla Smrža, Ph.D. Uviedol som všetky literárne premene, publikácie a ďalšie zdroje, z ktorých som čerpal.

.....
Juraj Samuel Saloň
10. mája 2021

Poděkování

Týmto by som sa chcel veľmi poděkovat pánovi doc. RNDr. Pavlovi Smržovi, Ph.D., za jeho ochotu, odbornú pomoc a obetený čas venovaný vedeniu tejto bakalárskej práce. Ďalej by som sa chcel poděkovat svojej rodine za podporu.

Obsah

1	Úvod	3
2	Systém pre prehľadávanie videí	5
2.1	Na čo slúži prehľadávanie videí?	5
2.2	Zloženie typického systému	6
2.3	Etické vplyvy a právne otázky	8
3	Rozpoznávanie objektov	9
3.1	Strojové videnie	9
3.2	Rozpoznávanie objektov pomocou neurónových sietí	10
3.3	Model You Only Look Once	14
3.4	Vyhodnocovanie modelov	15
3.4.1	Stredná priemerná presnosť	16
3.4.2	Skóre F1	17
4	Sledovanie objektov	19
4.1	Úvod do sledovania objektov	19
4.2	Algoritmus Deep Sort	21
5	Rozpoznávanie činností	22
5.1	Úvod do rozpoznávania akcií	22
5.2	Výzvy metód rozpoznávajúcich akcie	23
5.3	Prístup neurónových sietí k rozpoznávaniu akcií	24
5.4	Model SlowFast	25
5.5	Rozpoznávanie akcií pomocou kompozície sémantických informácií	27
6	Návrh a implementácia	30
6.1	Spracovanie dát	31
6.2	Model pre rozpoznávanie objektov	31
6.3	Model pre sledovanie objektov	33
6.4	Model pre rozpoznávanie akcií	35
6.5	Databáza	36
6.6	Definícia činností	37
7	Testovanie	40
8	Záver	46
	Literatúra	48

A Obsah priloženého pamäťového média	56
B Plagát	57

Kapitola 1

Úvod

V dnešnom svete, kedy sa výpočetná technika a rôzne kamerové systémy stali dostupnými aj pre bežnú verejnosť, vzniká konkurenčné súperenie medzi distribútormi bezpečnostných a sledovacích systémov. S bezpečnostnými systémami sa môžeme stretávať na každom kroku. Bežní ľudia ich používajú na zabezpečenie svojich pozemkov, polícia dokáže pomocou nich identifikovať páchateľa alebo sa používajú len na sledovanie dopravnej situácie v úsekokach, kde sa často vyskytujú dopravné nehody.

Kamerový systém sám o sebe nemá väčší význam, pokiaľ nikto nesleduje, čo systém zachytil. Človek však podlieha rôznym psychickým vplyvom, napr. únava a stres, a preto nie je možné, aby dokázal nepretržite sledovať udalosti na kamerových záznamoch s úplnou sústredenosťou. Niektorí ľudia využívajú kamerové bezpečnostné systémy len na monitorovanie objektov a záznamy sledujú, len keď vedia o nejakej situácii, ktorá nastala. Čo však v prípade, že o nejakej situácii nevedia? Mnoho prípadov, napríklad krádeží, preto ostáva nevyriešených. Z tohto dôvodu začali byť vyvíjané rôzne automatické systémy schopné vykonávať prehľadávaciu činnosť za človeka.

Velký vplyv na vývoj systémov pre rozpoznávanie činností vo videu majú práve spracovávanie dát a technológia internetu. Postupné zdokonalovanie týchto technológií umožnilo upravovať, ukladať a presúvať veľké množstvo dát rôznych typov. Väčšina sledovacích systémov je založená na metódach strojového učenia práve pre jednoduchú dostupnosť obrovských objemov dát, ktoré sú potrebné pre trénovanie algoritmov strojového učenia, v poslednej dobe predovšetkým neurónových sietí.

Metódy strojového učenia potrebujú pre rozpoznávanie určitého typu akcie alebo činnosti veľké množstvo anotovaných dát, ktoré však nie sú pre väčšinu takých činností a situácií k dispozícii. Používané systémy často nedokážu rozpoznávať niektoré typy zložených akcií, pri ktorých je potreba súčasne kombinovať viaceré objekty alebo informácie. Takéto udalosti môžu byť popísané pomocou vzťahov medzi objektami. Príkladom udalosti môže byť „nebezpečné prechádzanie chodcov na červenú“, kedy je potrebné pre rozpoznanie takejto situácie poznáť polohu človeka a farbu na semafore. Informácie tohto typu je už dnes možné rozpoznávať s vysokou presnosťou pomocou rôznych metód a modelov neurónových sietí.

Cieľom tejto práce je vytvorenie systému, ktorý spojí viaceré modely pre získanie jednoduchých informácií, uloží tieto informácie do databázy a využije ich pre vyhľadávanie činností, definovaných pomocou vzťahov medzi ukladanými informáciami. Systém získava úseky videa z kamerového záznamu, ktoré pravdepodobne zaznamenávajú vyhľadávanú činnosť. Vzhľadom k miere neistoty ako pri získavaní jednotlivých informácií, tak pri definovaní vzťahov a vyhľadávaní informácií v databáze, je nutné naviac uvažovať o tejto úlohe ako o

usporiadanie jednotlivých úsekov videa na základe pravdepodobnosti toho, že ide o požadovanú akciu.

Text práce je rozdelený do dvoch častí. Prvú časť otvára kapitola 2, ktorá objasňuje problematiku systémov pre rozpoznávanie ľudského správania a činností. Nasledujúce kapitoly potom rozoberajú postupne jednotlivé základy a techniky pre rozpoznávanie objektov 3, sledovanie objektov 4 a rozpoznávanie akcií 5, využité v tejto práci.

Druhý celok sa zameriava na porovnanie a výber jednotlivých modelov pre získavanie informácií a spracovávanie vizuálnych dát. Obsahuje dve kapitoly, z ktorých prvá popisuje odôvodnenie výberu modelov neurónových sietí pre jednotlivé moduly systému a implementačné detaily (kapitola 6). Posledná kapitola 7 ukazuje pohľad na výsledné testovanie definovaných činností a vyhodnotenie schopnosti rozpoznať ľudské činnosti. Záver a smery možného rozšírenia vytvoreného systému sú diskutované v kapitole 8.

Kapitola 2

Systém pre prehľadávanie videí

Zameraním tejto práce je práve vytvorenie časti systému, ktorá je schopná rozpoznávania činností a akcií ľudí. Táto schopnosť je najdôležitejším prvkom systémov pre prehľadávanie videa, a preto je vhodné sa najprv zoznámiť s konceptom týchto programov (sekcia 2.1). Každý takýto systém sa skladá z niekoľkých častí, ktoré sú jednoducho popísané v sekcií 2.2. Následne sú v sekcií 2.3 spomenuté niektoré etické a právne otázky, ktorým takéto systémy podliehajú.

2.1 Na čo slúži prehľadávanie videí?

V poslednej dobe sa vládne agentúry, podniky a dokonca aj školy obracajú k video sledovaniu ako prostriedku na zvýšenie verejnej bezpečnosti. Rozširovanie lacných kamier a dostupnosť vysoko-rýchlosných širokopásmových bezdrôtových sietí sa nasadením veľkého počtu týchto kamier na kontrolu bezpečnosti stalo ekonomicky a technicky uskutočnitelným. S rastúcim počtom sledovacích kamier vo vnútorných aj vonkajších priestoroch rastie dopyt po inteligentných systémoch, ktoré rozpoznávajú rôzne udalosti. Najnovšie práce sa pokúšajú integrovať schopnosti počítačového videnia, spracovania obrazu a umelej inteligencie do aplikácií na sledovanie videa. Na zaistenie bezpečnosti ľudí je navrhnutých niekoľko systémov. Je dôležité presne opísat video-obsah a umožniť organizovanie a vyhľadávanie potenciálnych videí, aby bolo možné zistiť a analyzovať súvisiace udalosti sledovania [37, 67, 55, 42].



Obr. 2.1: Príklad situácie, ktorú je potrebné rozpoznávať z dôvodu ohrozenia ľudských životov (prevzaté z [59])

Široká dostupnosť údajov z kamier, ktoré sú inštalované napríklad v obytných štvrtiach, priemyselných podnikoch, vzdelávacích inštitúciách a obchodných firmách patria k súkromnému údajom. Naopak videozáznamy z kamier umiestnených na verejných miestach, ako sú centrá miest, verejné dopravné prostriedky a náboženské miesta, patria k verejným údajom. Údaje zo systémov pre sledovanie videa majú v dnešnom svete veľký spoločenský význam. Hlavnými dôvodmi, prečo sa zaoberať práve touto témom systémov pre prehľadávanie videí, sú:

- umožňujú nepretržité sledovanie videí, ktoré je pre ľudí veľmi náročné a únavné,
- inteligentný systém pre prehľadávanie videí je významným nástrojom pre ľudí, ktorí namiesto nich robí namáhavú úlohu,
- potreba presnej identifikácie a rozpoznávania ľudského správania a činností,
- potreba vylepšenia niektorých odvetví, akým je napríklad analýza správania ľudí v dave,
- potreba dokázať rozpoznávať rôzne situácie ohrozujúce život v aktuálne prebiehajúcom čase,
- predpovedanie určitého pohybu, abnormálneho správania alebo násilia je veľmi užitočné v život ohrozujúcich situáciach (obrázok 2.1),
- obrovská dostupnosť jednotlivých dát v podobe videozáznamov.

Viac o problematike týkajúcej sa prehľadávania videí a cieľoch odvetvia pojednávajú články [58] a [40].

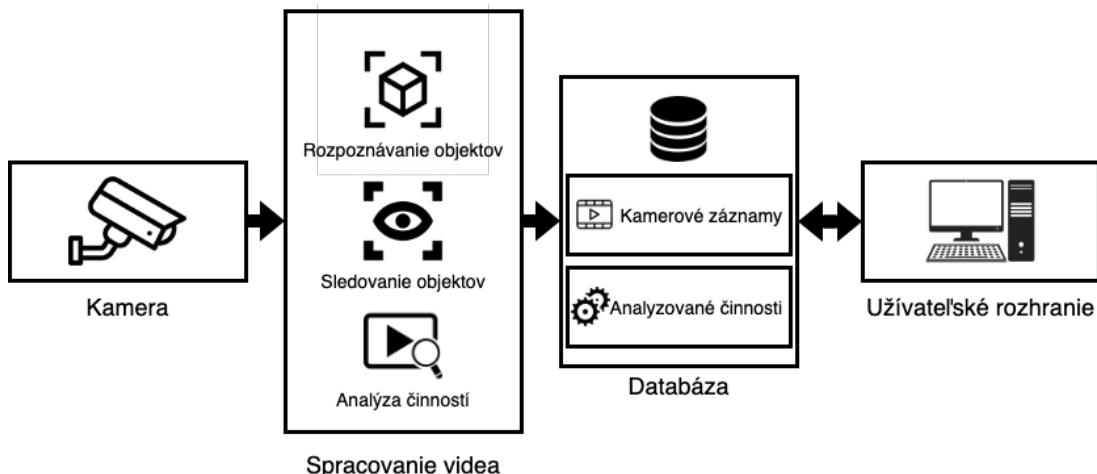
2.2 Zloženie typického systému

V dnešnej dobe sú videozáznamy, zachytené zo strategicky umiestnených kamier, hlavným prvkom každého monitorovacieho systému. Systémy používané pre rôzne účely sa môžu v rôznych formách lísiť. Napríklad systém pre udržanie bezpečnosti na letisku vyžaduje striktnejšie pravidlá pre vyhodnocovanie nejakých podozrivých činností a posúdenie v aktuálne prebiehajúcom čase, či ide o hrozbu teroristického útoku alebo abnormálne správanie človeka, aby sa dokázalo predísť stratám na životoch. Naopak systém vytvorený pre kontrolu dochádzky žiakov do školy nepotrebuje rozpoznávať jednotlivých žiakov v reálnom čase, no stačí, keď vykoná svoju aktivitu raz denne na videozáznamoch získaných počas dňa. Aj keď sú systémy používané na rôzne účely a disponujú rôznymi rozdielnymi funkcionalitymi, spoločná architektúra klasifikácie pozostáva z rovnakých častí alebo krokov [13, 4] (obrázok 2.2):

- kamera,
- rozpoznanie objektov,
- sledovanie objektov,
- analýza činností,
- databáza,

- užívateľské rozhranie.

Kamera. Kamera je základným prvkom každého systému pre prehľadávanie videa. Pomocou kamery sú snímané jednotlivé oblasti reálneho sveta, nad ktorými je vykonávaná analýza a z ktorého sú rozpoznávané činnosti ľudí. Jej kvalita, rozlíšenie a snímkovanie za sekundu, má veľký vplyv na analýzu. Čím väčšie rozlíšenie kamera používa, tým je systém schopný rozpoznať viac informácií z jednotlivých snímkov.



Obr. 2.2: Zloženie typického systému pre prehľadávanie videí

Rozpoznávanie objektov. Rozpoznanie objektov je prvým stupňom analýzy videa vo väčšine sledovacích systémov a slúži ako prostriedok na zameranie pozornosti a klasifikáciu objektu. Problém rozpoznávania objektov sa rieši pomocou rôznych metód a algoritmov, akými sú napríklad aj neurónové siete. Informácie získané pomocou rozpoznávania objektov sú často potrebné pre fungovanie ďalších častí systému [42].

Sledovanie objektov. Účelom sledovania je určiť časopriestorové informácie o každom cieli prítomnom na scéne. Vizuálny pohyb cieľov je v porovnaní s ich priestorovými rozsahmi vždy malý. Sledovanie tiež často interaguje s rozpoznávaním objektov. Na sledovanie objektov vo video-sekvenciach sa tiež používa niekoľko techník, aby sa vyrovnali s viacerými interagujúcimi cieľmi. Analýza ľudského pohybu pomáha pri riešení problémov vo vnútorných monitorovacích aplikáciach [40].

Analýza činností. Analýza činností je prvkom systému, pomocou ktorého sú získané výsledné, užívateľom požadované, hľadané činnosti. Analýza funguje na princípe, kedy sú informácie získané spracovávaním videa využité pre definíciu činností alebo správania človeka. Typicky v prípade, že daná činnosť nastane alebo sa správanie človeka vychýli od definovaného normálneho správania, systém upozorní operátora systému [28].

Databáza. Databáza slúži pre ukladanie dát, akými sú videozáznamy z kamier alebo získané informácie pomocou spracovávania videa. Na základe využitých druhov databáz je potom možná rôzna práca s uloženými dátami. Niektoré systémy využívajú databázy, pomocou ktorých dokážu pracovať s časovo-priestorovými údajmi, a tým umožňujú jednoduché získavanie a vyhľadávanie vyskytnutých situácií vo videu.

Užívateľské rozhranie. Na manipuláciu s jednotlivými videami a kamerovými záznamami slúži užívateľské rozhranie. V závislosti na druhu systému pre prehľadávanie videa, konečný užívateľ vkladá do systému videozáznam alebo pozoruje scény získané pomocou kamier v reálnom čase. Užívateľské rozhranie sa sústredí na čo najintuitívnejšie ovládanie

a manipuláciu s celým systémom ako s celkom a na prezentáciu jednotlivých výsledkov získaných pomocou analýzy a spracovania videozáznamov.

2.3 Etické vplyvy a právne otázky

Pri inteligentných bezpečnostných systémoch sa naskytuje otázka, kde a kedy je možné takýto systém nasadiť. Viditeľné kamerové systémy ovplyvňujú a normalizujú ľudské správanie. Pokiaľ človek vidí kameru, je väčšia pravdepodobnosť, že nebude chodiť na zakázané miesta alebo vykonávať nežiadúce činy. Pre nasadenie takejto technológie je potrebné zvážiť okolnosti a umiestnenie, v ktorých bude fungovať. V prostrediaciach, ako je letisko, sa výrazne lísi forma normálneho správania sa od iných miest, ako napríklad štadión. Ďalšou dôležitou úlohou inteligentných prehľadávacích systémov je nájsť rovnováhu medzi správaním považovaným za abnormálne a normálne, a tým sa vyhnúť diskriminácii alebo spôsobiť ľuďom nepríjemnosti. Napríklad pohybový model zdravého človeka sa výrazne lísi od modelu človeka na vozíku, čo môže spôsobiť chybnú identifikáciu abnormálneho správania (obrázok 2.3), a tým vysvetliť nedorozumenie. Preto si tvorba takýchto systémov vyžaduje spoluprácu medzi špecializovanými návrhármi systému a etikmi [20].



Obr. 2.3: Príklad pohybu vykonávaného človekom na vozíku, ktorý môže byť chybne rozpoznaný ako abnormálne správanie²

Súčasné schopnosti systémov pre prehľadávanie videa sú pre orgány činné v trestnom konaní na celom svete veľmi atraktívne a tieto systémy vidia ako účinný mechanizmus boja proti bezpečnostným hrozbám. Niektorí kritici však tvrdia, že všadeprítomné bezpečnostné kamery predstavujú hrozbu pre mnohé demokratické práva bežných ľudí. Z dôvodu rozsiahlych kamerových systémov sú v platnosti určité ústavou chránené hodnoty. Prvým takýmto právom je v právo na anonymitu, ktoré úzko súvisí so súkromím. Mnoho ľudí očakáva, že zostanú v anonymite na verejných miestach, ako sú napríklad kliniky pre neplodnosť alebo psychiatrické ordinácie. Po druhé, je ohrozené demokratické právo ľudí slobodne vyjadrovať svoje myšlienky a slobodne sa združovať a zdieľať ich. Ľudia by sa možno necítili pohodlne, keby vyjadrili svoje názory alebo sa zúčastnili protestov proti vládnej politike s vedomím, že ich pre túto činnosť môžu neskôr identifikovať [46].

²prevzaté z <https://www.friendshipcircle.org/blog/2012/06/05/air-travelers-with-disabilities-here-are-your-rights/>

Kapitola 3

Rozpoznávanie objektov

Schopnosť rozpoznať objekty je kľúčovou schopnosťou celého systému. Bez tejto schopnosti by ostatné časti systému nedokázali fungovať. Pomocou rozpoznávania objektov sú taktiež získavané informácie, akými sú napríklad poloha objektu alebo jeho trieda, ktoré sú veľmi dôležité pri analýze činností z videa. V tejto kapitole je postupne rozobraná problematika strojového videnia (sekcia 3.1) a rozpoznávanie objektov pomocou neurónových sietí (sekcia 3.2). Následne je v sekcií 3.3 vysvetlený princíp fungovania konkrétneho modelu konvolučnej neurónovej siete použitej v tejto práci a v sekcií 3.4 je ukázaný spôsob testovania takýchto modelov.

3.1 Strojové videnie

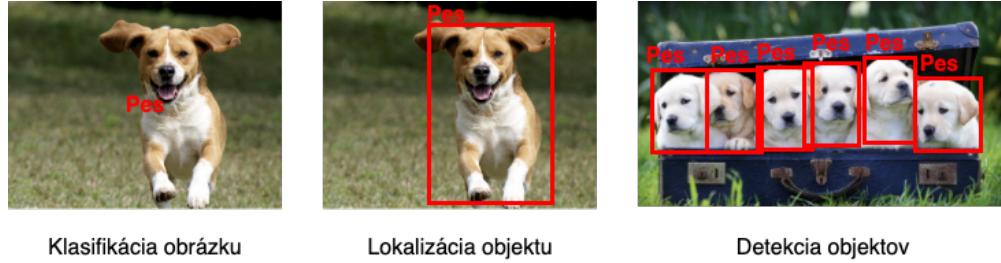
Strojové videnie má dvojaký cieľ. Z biologicko-vedeckého hľadiska je jeho cieľom nahrať ľudský vizuálny systém pomocou výpočtových modelov. Z hľadiska inžinierstva je cieľom strojového videnia vytvárať autonómne systémy, ktoré môžu vykonávať niektoré úlohy, zhodné s úlohami ľudského vizuálneho systému a dokonca ich v istých ohľadoch aj prekonať [23]. Mnohé vizuálne úlohy úzko súvisia s časovými a 3D informáciami extrahovanými z časovo-variabilných 2D dát získaných z jednej alebo viacerých kamier pre všeobecné porozumenie dynamických scén.

Vlastnosti a charakteristiky ľudského systému pre spracovávanie vizuálnych vnemov sú často inšpiráciou pre inžinierov a vedcov pri vytváraní strojových vizuálnych systémov. Ľudský vizuálny systém ponúka jednoduchý pohľad na to, ako strojové videnie funguje.

Mozog rozdeľuje vizuálne vnemy do kanálov, ktoré posielajú rôzne druhy informácií ďalej do mozgu. Tieto druhy informácií v mozgu prechádzajú do systému pozornosti v závislosti na vykonávanej úlohe. Systém je zodpovedný za nájdenie oblastí z obrazov pre ďalšie podrobnejšie preskúmanie a zároveň za potlačenie ostatných nezaujímavých oblastí [25]. Pomocou rozsiahlej asociácie vstupov zo senzorov vizuálneho systému a ostatných zmyslov je mozog schopný vykresliť určité informácie alebo kontext na základe dlhoročne získavaných skúseností počas ľudského života.

Strojové videnie je štúdia zaobrajúca sa transformáciou dát z obrázkov a video sekvenčí, ktorá umožňuje počítačom pochopit a interpretovať vizuálne informácie [5]. Takéto transformácie sú vytvárané s presným zámerom dosiahnutia vopred určených cieľov. Ide napríklad o získavanie kontextových informácií zo vstupných dát, akými sú „kamera sa nachádza na aute“ alebo „laser, zistujúci vzdialenosť, zachytil objekt vzdialenosť jeden meter“ [25].

Odvetvie strojového videnia sa zaobráva v podstate všetkým, čo ľudia vidia a vnímajú pomocou vizuálnych vnemov. Ľudia takéto úlohy vykonávajú podvedome na základe získaných skúseností, avšak pre počítače sú tieto úlohy veľkou výzvou. K úlohám strojového videnia sa zaraďuje aj schopnosť počítača rozpoznávať objekty.

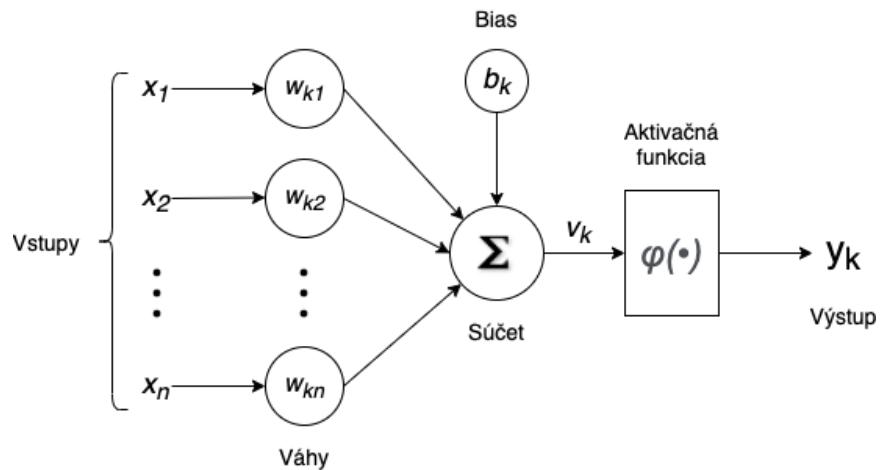


Obr. 3.1: Znázornenie rozdielu medzi klasifikáciou obrázku, lokalizáciou objektu a detekciou objektov

Rozpoznávanie objektov je základný pojem pre popis kolekcie súvisiacich úloh strojového videnia zahrňajúcich identifikáciu objektov z digitálnych obrázkov. Táto kolekcia sa skladá z klasifikácie obrázku, lokalizácie objektu a detekcie objektov (obrázok 3.1). Klasifikácia obrázku je proces, pri ktorom počítač berie daný obrázok a zaraďuje ho do určitých kategórií [50]. Pomocou lokalizácie objektu počítač získava rovnaké informácie o kategórií ako pri klasifikácii obrázku, avšak naviac získava približnú pozíciu a rozmeru objektu vyskytujúceho sa na danom obrázku. Nakoniec detekcia objektov slúži pre klasifikáciu a lokalizáciu viacerých objektov vyskytujúcich sa na snímku.

3.2 Rozpoznávanie objektov pomocou neurónových sietí

V posledných rokoch je vo vedeckých kruhoch obrovský záujem o neurónové siete, a to najmä pre svoje schopnosti vykonávať náročné výpočetné úlohy. Uzly a umelé neuróny si v neurónových sietiach berú inšpiráciu zo skutočných neurónových buniek v živých organizmoch. Zároveň distribúcia váh medzi neurónmi pripravuje synapsie, ktoré prepájajú jednotlivé bunky [62].



Obr. 3.2: Znázornenie jednej stavebnej bunky neurónovej siete – neurónu

Neurón. Vstupom neurónu je forma multi-rozmerného vektora vložená do vstupnej vrstvy, ktorá ho distribuuje hlbšie do skrytých vrstiev neurónovej siete. Z obrázku 3.2 je viďet, že do neurónu prichádzajú vstupné dátá v podobe hodnôt x_1, x_2, \dots, x_n . Každá hodnota vstupných dát je vynásobená príslušnou váhou ($w_{k1}, w_{k2}, \dots, w_{kn}$) a spočítaná s ostatnými násobkami. K výslednému súčtu je pripočítaný bias (b_k). Daná hodnota je poslaná do aktivačnej funkcie, ktorá rozhodne o aktivite neurónu, čiže o hodnote výstupu (y_k) [69]. Tento proces je vyjadrený pomocou vzorca:

$$y_k = \varphi\left(\sum_{i=1}^n x_i w_{ki} + b_k\right) \quad (3.1)$$

Aktivačná funkcia. Aktivačná funkcia transformuje hodnotu získanú zo súčtov násobkov váh a hodnôt vstupných dát. Výsledkom transformácie je hodnota, ktorá je použitá ako vstup do ďalšej vrstvy siete. Pokiaľ neurónová sieť nepoužíva aktivačnú funkciu, výstupným signálom siete je jednoduchá lineárna funkcia. Táto funkcia je ľahká pre vypočítanie, no jej komplexnosť je obmedzená. Modely neurónových sietí bez aktivačných funkcií nemajú potrebné schopnosti vykonávať úlohy, akými sú modelovanie zložitých typov dát (obrázky, videá alebo zvukové nahrávky). Medzi najznámejšie aktivačné funkcie patria napríklad Sigmoid alebo ReLU.

Sigmoid je typ nelineárnej funkcie, ktorá transformuje hodnoty v rozmedzí 0 až 1. Je vyjadrená pomocou vzorca:

$$f(x) = 1/e^{-x} \quad (3.2)$$

Zároveň však táto funkcia nie je symetrická okolo hodnoty 0, z čoho vyplýva, že hodnota každého neurónu bude kladná. Tento problém môže byť vyriešený pomocou úpravy funkcie, a preto je známejšia jej úprava $f(x) = 1/(1 - e^{-x})$.

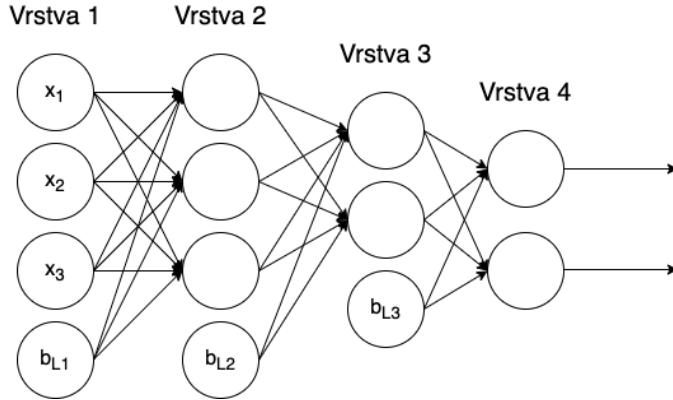
V poslednom čase je veľmi používaná aj aktivačná funkcia *Rectified Liner Unit* pod skratkou ReLU. Pri používaní tejto funkcie bývajú aktivované len niektoré neuróny [54]. Pokiaľ je vstupom hodnota nižšia ako 0, neurón je neaktívny. Jej definícia je:

$$f(x) = \max(0, x) \quad (3.3)$$

Vrstvy. Každá neurónová sieť obsahuje minimálne tri vrstvy. Vstupnú, výstupnú a skrytú vrstvu. Vrstvy sú zvyčajne usporiadane do stĺpcov, ako je možné vidieť na obrázku 3.3. Každá vrstva neurónov n je prepojená spojeniami s vrstvami $n - 1$ a $n + 1$. Ku každej vrstve je pripojený bias b_{Ln} upravujúci hodnotu. Pokiaľ sieť obsahuje viacero ako jednu skrytú vrstvu, nazýva sa *Hlboká Neurónová Sieť* [34].

Učenie. Po vytvorení neurónovej siete nie je hned možné, aby dávala pravdivé výsledky, a preto využíva proces učenia nazývaný učenie s učiteľom. Pri tomto procese sa neurónová sieť učí pomocou predznačených dát, ktorých výsledok je vopred známy [41]. Pre každú jednotku trénovacích dát existuje jedna alebo viacero vopred známych výsledných hodnôt. Neurónová sieť môže byť učená pomocou rôznych techník, z ktorých najznámejšou technikou učenia je Spätná Propagácia chyby (v angl. Back-Propagation) v kombinácii so Zostupom Gradientu (v angl. Gradient Descent).

Optimalizácia chyby. Pri učení dochádza k získaniu výstupnej hodnoty na základe výpočtu medzi vstupnými dátami a váhami neurónov. Po získaní výstupu však treba určiť rozdiel medzi výstupnou a očakávanou hodnotou a na základe toho upraviť jednotlivé parametre siete.



Obr. 3.3: Ukážka prepojenia neurónov v neurónovej sieti

Po získaní výstupu z trénovacích dát je ďalším krokom upravenie jednotlivých váh neurónov, ktoré zabezpečuje spätná propagácia chyby. Algoritmus sa späť pozera na jednotlivé váhy, ktoré najviac ovplyvnili výstupnú hodnotu a upraví ich pomocou parametru rýchlosťi učenia (v angl. learning rate). Úprava váh neurónov je definovaná pomocou vzorca:

$$w_{ij}(t+1) = w_{ij}(t) - \varepsilon \frac{\partial E}{\partial w_{ij}}(t) \quad (3.4)$$

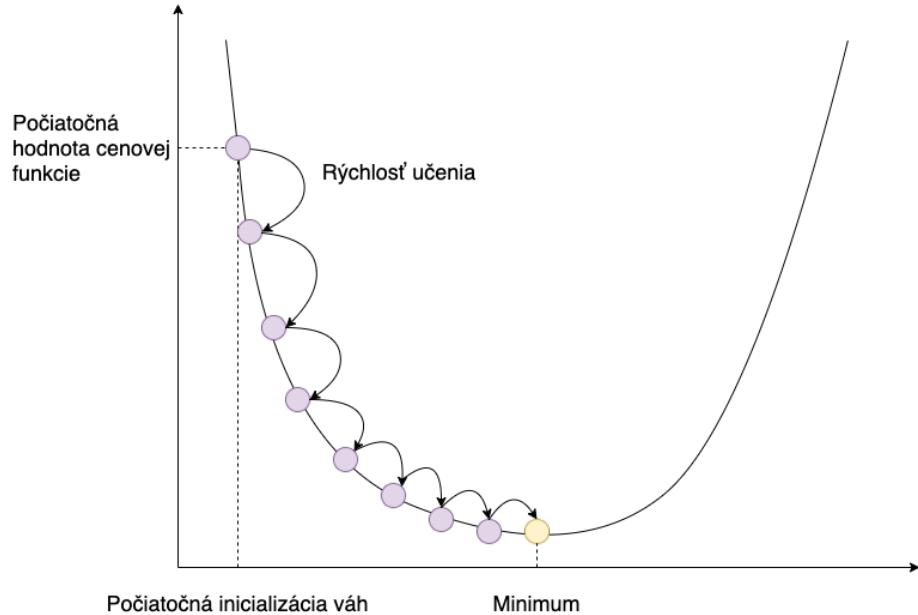
kde $w_{ij}(t)$ predstavuje aktuálnu hodnotu váhy neurónu, ε je parameter rýchlosťi učenia a $\frac{\partial E}{\partial w_{ij}}(t)$ je pomer vyjadrujúci hodnotu zostupu gradientu [49]. Na obrázku 3.4 je možné vidieť, ako zmeny váh neurónov ovplyvňujú klesanie cenovej funkcie.

O získanie miery chybovosti siete sa stará algoritmus Zostupu Gradientu. Zostup Gradientu je možné si predstaviť ako zostup z kopca, kedy turista hľadá najnižší bod v údolí. Tento algoritmus vyžaduje cenovú funkciu, určujúcu, aká je aktuálna veľkosť chyby siete. Vypočítaním rozdielu cenovej funkcie medzi výstupnou hodnotou siete a očakávanou hodnotou určí, ako majú byť jednotlivé váhy neurónov upravené, aby sa v ďalšom cykle učenia výsledky priblížili požadovanému výstupu [64].

Konvolučné neurónové siete. Konvolučná neurónová siet je rozšírením neurónovej siete, ktorá slúži pre spracovávanie informácií z vizuálnych dát. Tento druh siete je schopný učenia popisných vlastností odolných voči posunu a rotácií objektov. Popisné vlastnosti sú druh filtra, ktorý môže byť použitý na celý obrázok [45]. Konvolučné neurónové siete majú schopnosť hierarchického učenia popisných vlastností, ktorých výskum ukazuje, že extra-hované vlastnosti majú silnejšiu schopnosť popisu a generalizácie ako ručne vypočítavané vlastnosti (v angl. hand-crafted feature) [71].

V štandardnej štruktúre konvolučnej neurónovej siete sa zvyčajne striedajú konvolučné vrstvy nasledované pooling vrstvou a nelineárnu vrstvou. V závislosti na hĺbke siete sa táto štruktúra opakuje. Za vrstvami nakoniec nasleduje jedna alebo viacej plne prepojených vrstiev [53, 71].

Konvolučná vrstva je prvou vrstvou, ktorá extrahuje popisnú vlastnosť obrázka. Vy- užíva matematickú operáciu konvolúcia, ktorej vstupmi sú obrázok a filter reprezentované pomocou matice čísel. Konvolúcia pomocou filtrov vykonáva akcie, akými sú napríklad rozpoznávanie hrán, rozmažanie obrazu alebo zaostrenie. Vstupom a výstupom každej konvolučnej vrstvy je množina polí, inak nazývaných aj mapy popisných vlastností (v angl. fe-

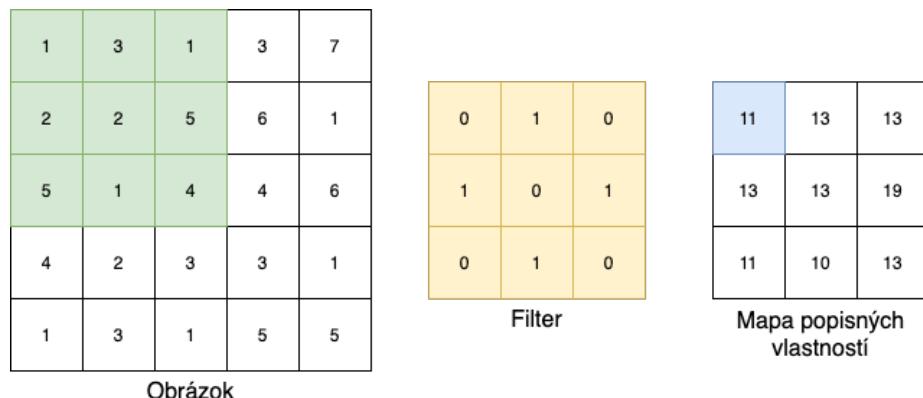


Obr. 3.4: Príklad grafu znázorňujúceho klesanie cenovej funkcie každým ďalším cyklom učenia

ature maps), ktoré predstavujú určitú oblasť obrázku. Konvolúcia je definovaná pomocou matematického vzorca ako [44]:

$$y[i, j] = \sum_{m < L} \sum_{n < H} h[m, n] \cdot x[i - m, j - n] \quad (3.5)$$

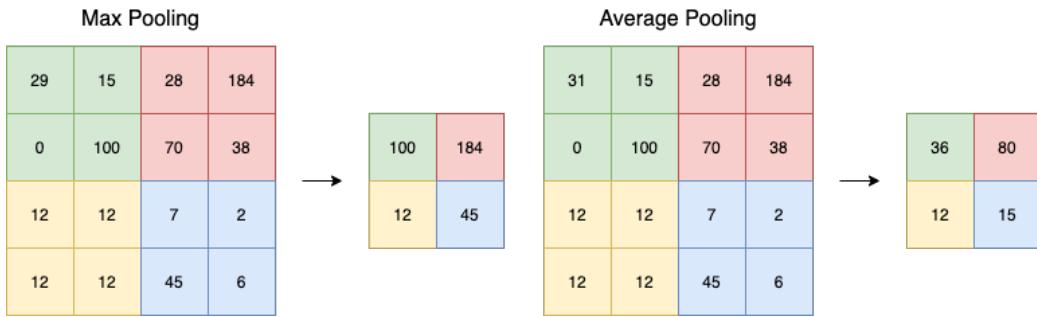
kde x reprezentuje maticu vstupného obrázku alebo oblasti obrázku, h predstavuje filter, H a L sú výška a šírka vstupného obrázka a y je výslednou mapou popisnej vlastnosti (obrázok 3.5). Získaná mapa popisných vlastností reprezentuje určitú vlastnosť extrahovanú na každej lokácii vstupného regiónu obrázka, ktorá je opäť reprezentovaná ako matica čísel [29].



Obr. 3.5: Príklad vypočítania konvolúcie medzi obrázkom a filtrom

Hlavnou ideou pooling vrstvy je znižovanie rozlíšenia a redukovanie zložitosti pre ďalšie vrstvy. S každou mapou popisných vlastností zaobchádzá samostatne. Táto vrstva rozdeľuje

vstupnú popisnú vlastnosť na menšie regióny, nad ktorými vypočítava hodnoty v závislosti na typie pooling vrstvy [1]. Najbežnejšími používanými typmi pooling vrstiev sú *Max-Pooling*, ktorá vyberá maximálnu hodnotu z regiónu a *Average-Pooling*, ktorá počíta priemernú hodnotu zo všetkých hodnôt v regióne. Jednotlivé susedné regióny sú od seba vzdialé minimálne s veľkosťou kroku vyššou ako 1. Na obrázku 3.6 je možné vidieť jednoduchý príklad obidvoch výpočtov týchto vrstiev nad vyznačenými regiónmi s veľkosťou kroku 2. Výsledkom je zredukovanie rozlíšenia výstupnej mapy popisných vlastností [29].



Obr. 3.6: Znázormenie výpočtu pomocou Max-Pooling a Average-Pooling vrstvy

Poslednou vrstvou konvolučnej siete je jedna alebo viacero plne prepojených vrstiev. Vstupom pre túto vrstvu je vektor čísel získaný z poslednej pooling vrstvy pomocou sploštenia získaných máp popisných vlastností. Počet uzlov poslednej plne prepojenej vrstvy je rovný počtu výsledných tried, ktoré je neurónová sieť schopná rozpoznať. Úlohou týchto vrstiev je získanie konečného výsledku v podobe klasifikácie obrázku alebo objektov na obrázku [60].

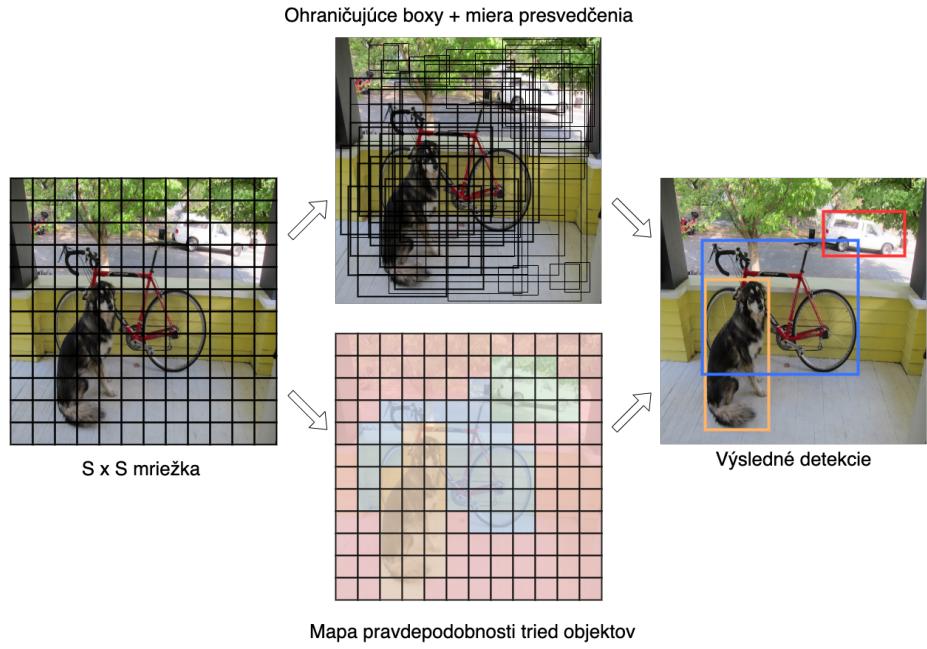
3.3 Model You Only Look Once

Jedným z modelov využívajúcich konvolučné neurónové siete je model s názvom *You Only Look Once* (dalej len YOLO). Tento model sa v dnešnej dobe vyskytuje v niekoľkých verziách, kedy každá verzia priniesla nejaké vylepšenia v porovnaní s predchádzajúcimi. Model svoju povest získal najmä pre schopnosť rozpoznať objekty v reálnom čase, zatiaľ čo jeho presnosť je porovnateľná s najpresnejšími modelmi. Na základe týchto skutočností bol model YOLO použitý aj v tejto práci.

Modely pre rozpoznávanie objektov sú rozdeľované na dve skupiny:

- jednofázové modely,
- dvojfázové modely.

Model *YOLO* patrí do skupiny jednofázových modelov, ktoré sa vyznačujú schopnosťou rozpoznať objekty pomocou jednej konvolučnej siete. Táto konvolučná siet dokáže navrhnúť regióny, v ktorých sa môže nachádzať objekt a zároveň aj určiť o aký objekt ide. Dvojfázové modely využívajú viaceré neurónové siete, ktoré navrhujú a zarovnávajú zaujmavé regióny obrázka pred výslednou detekciou objektov. Výsledkom takejto architektúry sú väčšie zložitejšie a komplexnejšie neurónové siete s vysokou presnosťou, avšak s nedostatočnou rýchlosťou rozpoznávania objektov. Viacej o rozdieloch jedno a dvojfázových



Obr. 3.7: Mriežka rozdeľujúca vstupný obrázok modelu YOLO

modelov pojednáva článok *MimicDet: Bridging the Gap Between One-Stage and Two-Stage Object Detection* [35].

YOLO rozdeľuje vstupný obrázok na mriežku veľkosti $S \times S$, ako je znázornené na obrázku 3.7. Pokiaľ do bunky mriežky spadá stred nejakého objektu, bunka je zodpovedná za jeho rozpoznanie. Úlohou týchto buniek je predpovedať ohraničujúce boxy a pravdepodobnosť výskytu objektu v boxe [47].

Každý ohraničujúci box je reprezentovaný pomocou piatich premenných (x, y, w, h, c) , kde (x, y) je stredový bod ohraničujúceho boxu v rámci bunky, (w, h) označuje šírku a výšku boxu vzhľadom na celý obrázok a nakoniec miera presvedčenia c vyjadruje ako presne daný box ohraničuje objekt [33].

Pre každú bunku je vypočítavaná trieda objektu, ktorá sa v nej s najvyššou pravdepodobnosťou nachádza. Tieto pravdepodobnosti sú obmedzené iba na bunky, v ktorých bol rozpoznaný objekt. Pravdepodobnosť je vypočítaná pre každý typ objektu, ktorý je neurónová sieť schopná rozpoznať. Bunka je potom priradená trieda s najväčšou mierou pravdepodobnosti bez ohľadu na to, koľko ohraničujúcich boxov obsahuje [47].

V mnohých prípadoch je jasné, do ktorej bunky objekt spadá a neurónová sieť navrhne pre takýto objekt iba jeden ohraničujúci box. Avšak v prípadoch kedy ide o veľký objekt alebo o objekt blízky hraniciam buniek, model preň navrhuje viacero boxov. Preto je potrebné odstrániť niektoré navrhnuté boxy. Pre tento problém je použitý algoritmus *Non-Max Suppression*. Výsledkom algoritmu je jeden ohraničujúci box, ktorý najlepšie vystihuje rozpoznaný objekt [10].

3.4 Vyhodnocovanie modelov

Pre spoľahlivé vyhodnotenie schopnosti úspešne rozpoznávať objekty je potrebné využiť vydohodnociaciu metriku, ktorá nám formálne určí, ako dobre dokáže model jednotlivé objekty

na obrázku identifikovať. V dnešnej dobe je najpoužívanejšou technikou v oblasti rozpoznavania objektov práve stredná priemerná presnosť (z angl. Mean Average Precision).

3.4.1 Stredná priemerná presnosť

Pre pochopenie strednej priemernej presnosti je nutné sa najprv zoznámiť s pojмami presnosť a úplnosť (z angl. precision a recall), ktoré slúžia pre vypočítanie konečnej hodnoty strednej priemernej presnosti.

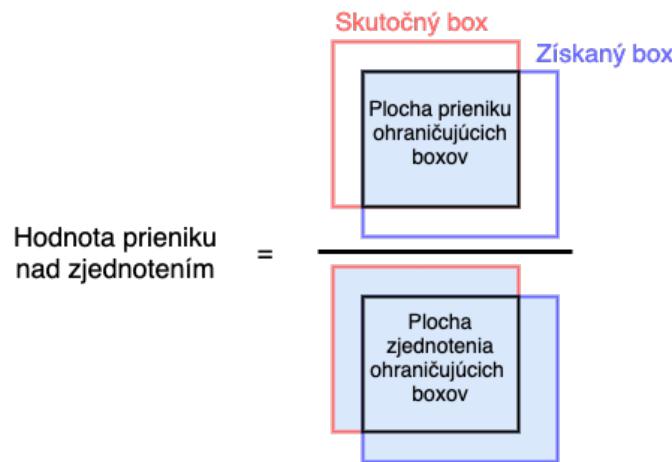
Ako z názvu vyplýva, v prípade rozpoznávania objektov presnosť udáva hodnotu, ktorá určuje ako presne model dokáže určovať pozitívne detekcie objektov. Táto hodnota je vydarená pomerom medzi pravdivo pozitívnymi detekciami (PP) objektov a všetkými pozitívnymi detekciami.

$$\text{Presnosť} = \frac{PP}{PP + FP} \quad (3.6)$$

Podobne ako presnosť, úplnosť určuje ako dobre dokáže model rozpoznávať objekty. Avšak v tomto prípade úplnosť udáva kolko bolo rozpoznaných pravdivo pozitívnych detekcií zo všetkých prípadov. Hodnota úplnosti je určená pomocou pomeru pravdivo pozitívnych detekcií a súčtu pravdivo pozitívnych detekcií s falošne negatívnymi detekciami (FN).

$$\text{Úplnosť} = \frac{PP}{PP + FN} \quad (3.7)$$

Pojmy pravdivá a nepravdivá detekcia určujú, s akou správnosťou dokázal model rozpoznať objekt. Pravdivé detekcie určujú, že sa model nezmýlil a správne určil, že sa objekt na danom obrázku nachádza (pravdivo pozitívna - PP) alebo sa nenachádza (pravdivo negatívna - PN). Naopak nepravdivé detekcie určujú pochybenie modelu, kedy model určil, že sa buď na obrázku daný objekt nachádza, aj keď tam nie je (falošne pozitívna - FP) alebo že sa objekt na obrázku nenachádza, no v skutočnosti sa tam vyskytuje (falošne negatívna - FN) [18].

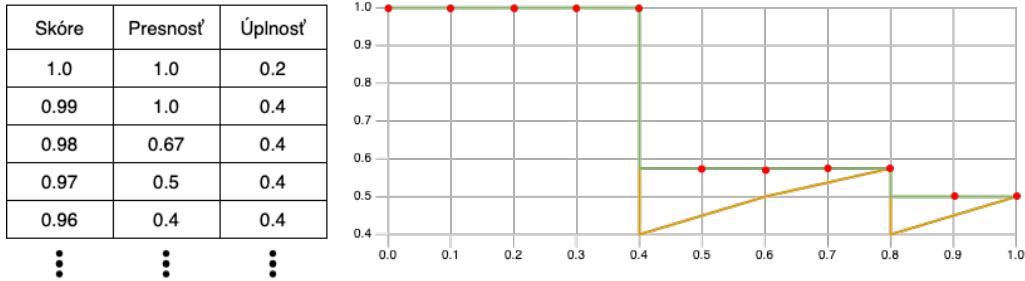


Obr. 3.8: Znázornenie výpočtu prieniku nad zjednotením, pomocou ohraňujúcich boxov

Pri rozpoznávaní objektov sa jednotlivé pravdivé a nepravdivé detekcie určujú z hodnoty prieniku nad zjednotením (z angl. intersection over union). Na obrázku 3.8 je možné vidieť, že hodnota prieniku nad zjednotením sa počíta pomocou skutočného manuálne navrhnutého

ahraničujúceho boxu a ohraničujúceho boxu získaného z modelu pre rozpoznávanie objektu. Pokiaľ táto hodnota prekročí určitú hranicu, výsledná detekcia je vyhodnotená ako pravdivo pozitívna.

Pre výpočet priemernej presnosti je potrebné zoradiť získané detekcie objektov v zostupnom poradí na základe hodnoty skóre, ktorú každej detekcií priraďuje model podľa toho, aký vierošodný je výsledok. Úspešným rozpoznaním je označená detekcia, ktorej hodnota prieniku nad zjednotením prekročí hranicu 0.5 [21]. Podľa týchto zoradených hodnôt je vypočítaná krivka presnosti a úplnosti, ktorú je možné vidieť na obrázku 3.9.



Obr. 3.9: Výpočet priemernej presnosti pomocou zoradených hodnôt presnosti a úplnosti

Výsledná priemerná presnosť pre určitý typ objektu je vypočítaná pomocou interpolácie krivky v jedenásťich bodoch stupnice úplnosti, ktorá môže byť vyjadrená vzorcom:

$$\text{Priemerná presnosť} = \frac{1}{11} \sum P(V_i) \quad i \in [0.0, 0.1, \dots, 1.0] \quad (3.8)$$

kde $P(V_i)$ predstavuje maximálnu hodnotu presnosti v určitej hranici úplnosti. Stredná priemerná presnosť je získaná po výpočte strednej presnosti pre každý typ objektu. Podrobnejšie vysvetlenie vyhodnocovania modelu pomocou strednej priemernej presnosti je objasnené v článku *A Survey on Performance Metrics for Object-Detection Algorithms* [43].

3.4.2 Skóre F1

F1 skóre alebo aj F skóre je metrika pre vyhodnocovanie binárnych systémov rozdeľujúci výsledky na „správne“ a „nesprávne“. Táto metrika je využívaná najmä pre hodnotenie systémov, ktoré získavajú informácie, akými sú napríklad vyhľadávače a mnoho druhov modelov strojového učenia.

F skóre je technika kombinácie presnosti (P) a úplnosti (U) modelu a je definovaná ako ich harmonický stred, kde vzorec pre výpočet harmonického stredu je nasledujúci [51]:

$$H = \frac{2}{\frac{1}{P} + \frac{1}{U}} \quad (3.9)$$

$$= \frac{2}{\frac{P+U}{PU}} \quad (3.10)$$

$$= \frac{2PU}{P+U} \quad (3.11)$$

Harmonický stred je viac intuitívnejšou technikou vyhodnotenia ako priemerná presnosť, pokiaľ sa jedná o vyhodnotenie nejakého systému, u ktorého je potrebné brať do úvahy ako správne, tak aj nesprávne výsledky.

V prípade kedy je testovaný systém pre prítomnosť výbušných látok, ktorý dosahuje presnosť s hodnotou 1, avšak úplnosť iba s hodnotou 0.2, je potrebné, aby bola úspešnosť tohto modelu veľmi nízka. Hodnota priemernej presnosti je v takomto prípade rovná najvyššej hodnote 1 a harmonický stred vykazuje hodnotu 0.333.

Metrika F skóre naviac dokáže naviac upravovať rovnováhu medzi presnosťou a úplnosťou na základe parametru β . Úplná definícia F skóre je vyjadrená pomocou vzorca ako:

$$F_\beta = \frac{(\beta^2 + 1)PU}{\beta^2 PU} \quad (0 \leq \beta \leq +\infty) \quad (3.12)$$

kde β je spomínaný parameter, ktorý v závislosti na jeho nastavení kontroluje rovnováhu medzi P a U. Pokiaľ je parameter $\beta = 1$, F_1 je rovný výpočtu harmonického stredu. V prípade, že $\beta > 1$ F skóre zohľadňuje viacej úplnosť a v opačnom prípade, kedy $\beta < 1$ je rovnováha obrátená a zohľadnená je viacej presnosť [9].

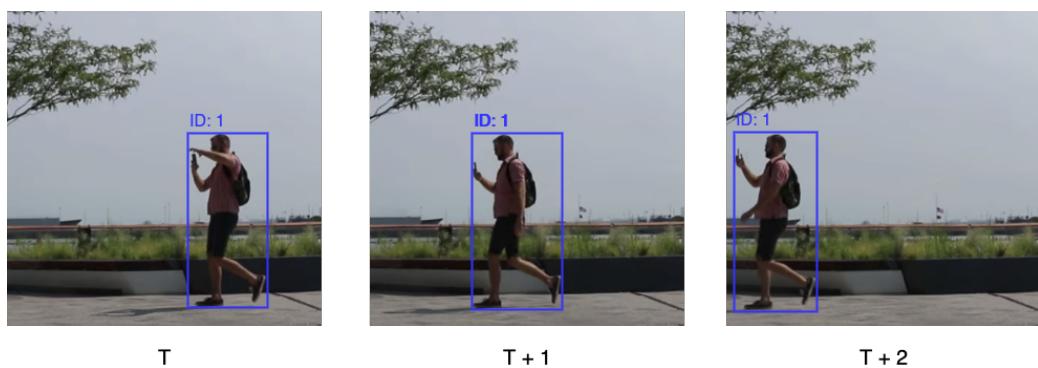
Kapitola 4

Sledovanie objektov

Súčasťou systému je schopnosť získavať úseky videí, v ktorých určitý človek alebo skupina vykonáva nejakú akciu. Preto musí byť systém schopný priradiť objektom určitý druh unikátneho identifikátoru po dobu výskytu vo videu. V nasledujúcej sekcii 4.1 bude postupne objasnené, akým spôsobom bývajú objekty sledované. Následne je v ďalšej sekcii 4.2 vysvetlený princíp fungovania konkrétneho algoritmu, ktorý je využívaný systémom implementovaným v tejto práci.

4.1 Úvod do sledovania objektov

Sledovanie objektov je disciplína počítačového videnia, ktorej úlohou je identifikovať trajektoriu objektov počas ich výskytu na videu alebo sekvencii obrázkov. Algoritmus pre sledovanie objektov priraduje objektu určitý druh identifikácie, akým je napríklad identifikačné číslo a nemení ho počas celého výskytu objektu. V závislosti na použitom algoritme pre sledovanie objektov môžu byť získané aj ďalšie dodatočné informácie, akými sú napríklad orientácia, plocha alebo tvar objektu [68]. Na obrázku 4.1 je možné vidieť príklad sledovania objektu s priradeným identifikačným číslom v rôznych časoch.



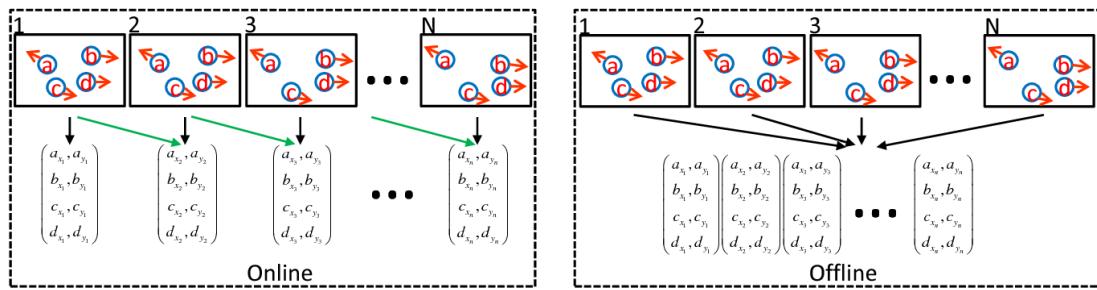
Obr. 4.1: Sledovanie objektu v rôznych časoch

Odvetvie sledovania objektov je rozdelené na dve hlavné skupiny podľa počtu sledovaných objektov na:

- sledovanie jedného objektu,
- sledovanie viacerých objektov.

Ako z názvu vyplýva, sledovanie jedného objektu slúži pre identifikáciu jedného objektu počas výskytu v sekvencii nadvážujúcich obrázkov. Pri tomto prístupe musí byť zvyčajne sledovaný objekt inicializovaný v prvom obrázku manuálne. Avšak z hľadiska použitia, napríklad v aplikáciach používaných pre identifikáciu správania ľudí alebo sledovanie dopravnej situácie, je viac zaujímavý prístup sledovania viacerých objektov [12].

Sledovanie objektov funguje na princípe extrahovania reprezentácie vzhľadu a pohybového modelu objektu. Reprezentácia objektu môže byť realizovaná pomocou rôznych techník, akými sú napríklad bodová reprezentácia, reprezentácia objektu pomocou jednoduchej kostry a podobne. V tejto práci je použitá reprezentácia objektu pomocou ohraničujúcich boxov získaných z modelu pre rozpoznávanie objektov, prebratom v kapitole 2. Algoritmus sledujúci objekty sa snaží čo najlepšie reprezentovať vzhľad a pohybový model daného objektu, ktorý porovnáva so získanými reprezentáciami objektov z predchádzajúceho obrázku [40]. Na základe miery podobnosti priraduje jednotlivým objektom identifikáciu podľa toho, či sa už na scéne videa vyskytli, alebo ide o nový objekt.



Obr. 4.2: Ukážka použitia informácií pomocou online a offline sledovania¹

Pre sledovanie objektov po dobu celého výskytu vo videu bývajú používané a porovnávané vzhľadové a pohybové informácie získané v rôznych okamihoch. Podľa týchto informácií v čase bývajú algoritmy rozdeľované na dve skupiny:

- online sledovanie,
- offline sledovanie.

Obrázok 4.2 znázorňuje, akým spôsobom využívajú získané informácie. *Online sledovanie* využíva pre porovnávanie objektov informácie o vzhľadovom a pohybovom modeli objektu získané v aktuálne spracovanom obrázku alebo informácie z predchádzajúcich obrázkov. Na rozdiel od *online sledovania*, *offline sledovanie* využíva všetky informácie získané buď v určitom časovom úseku, alebo z celej sekvenči obrázkov, z ktorej sa skladá video [36].

Najdôležitejšou funkciou algoritmov sledujúcich objekty je vhodná reprezentácia vzhľadu objektu. Veľkou výzvou pri získavaní tejto reprezentácie však bývajú problémy, s ktorými sa toto odvetvie stretáva už od vzniku. Najväčšími výzvami sú napríklad zmena osvetlenia prostredia, rotácia objektov alebo úplná oklúzia objektu [32]. Tieto problémy zamedzujú získaniu vhodnej reprezentácie vzhľadu, a tým znemožňujú sledovanie objektu po celú dobu výskytu pod rovnakým identifikačným číslom.

¹prevzaté z <https://d3i71xaburhd42.cloudfront.net/3dffacda086689c1bcb01a8dad4557a4e92b8205/8-Figure4-1.png>

4.2 Algoritmus Deep Sort

Deep Sort je jednoduchý algoritmus pre sledovanie objektov. Jeho najväčšou výhodou je schopnosť rozpoznávať objekty v reálnom čase. Využíva *Kalmanov filter* pre predikciu budúcej trasy objektu a *Hungarian algoritmus* pre riešenie problému priradenia trás k objektom.

Každé sledovanie objektu je reprezentované pomocou premenných $(u, v, \gamma, h, u', v', \gamma', h')$ kde (u, v) označuje stred ohraničujúceho boxu, γ označuje pomer strán a h je výška. Ostatné premenné vyjadrujú posledný sledovaný stav objektu [57].

Ku každej trase k je priradená hodnota a_k určujúca, akú dlhú dobu nebola trasa aktualizovaná. Pri predikcií Kalmanovým filtrom môže dôjsť k nepriradeniu k niektorým z trás, z dôvodu prekrytia alebo zmiznutia objektu zo scény. Pokial nie je aktualizácia trasy vykonaná po maximálnej dobe A_{max} , daná trasa je odstránená z množiny sledovaných trás.

Pri výskytu nového objektu je jeho trasa po dobu prvých troch predikcií označená ako predbežná. Prvé tri predikcie nemôžu byť prerušené, inak je daná trasa vymazaná a objekt je považovaný za taký, ktorý sa vo videu ešte neobjavil. Tento problém nastáva pomerne často, a preto *Deep Sort* využíva *Hungarian algoritmus*.

Hungarian algoritmus priraďuje predikcie porovnávaním pohybového a vzhľadového modelu objektu [65]. Porovnávanie pohybového modelu je vykonávané pomocou *vzdialenosťi Mahalanobis* medzi predikciou *Kalmanovho filtra* a novým ohraničujúcim boxom:

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i) \quad (4.1)$$

kde je projekcia i -tej trasy v priestore označovaná ako (y_i, S_i) a j -tý ohraničujúci box ako d_j . Pri výskytu problému s priradením predikcie je na základe strednej hodnoty pohybu medzi bodmi tras vypočítané, kde by sa mal daný objekt v aktuálnom čase nachádzať. Po každom výpočte je daný výsledok vložený do rozhodovacej logiky:

$$b_{i,j}^{(1)} = 1[d^{(1)}(i, j) \leq t^{(1)}] \quad (4.2)$$

ktorá je vyhodnotená na 1, pokial je i -tá trasa viditeľne priraditeľná k j -tému ohraničujúcemu boxu. Hranica ktorá nesmie byť prekročená má hodnotu $t^{(1)} = 9.4877$.

Vzdialenosť Mahalanobis je vhodnou technikou korekcie pohybu objektov, avšak predstavuje iba hrubý odhad pozície objektu.

Ďalšou technikou, ktorá zlepšuje riešenie problému priradenia je porovnávanie vzhľadového modelu pomocou kosínusovej vzdialenosťi. Pre každý ohraničujúci box d_j je pomocou jednoduchej neurónovej siete vypočítaný vzhľadový popis r_j . Ku každej trase k je priradená galéria $R_k = \{r_k^{(i)}\}_{i=1}^{L_k}$ posledných $L_k = 100$ vzhľadových popisov. Následne je počítaná najmenšia kosínusová vzdialenosť medzi i -tou trasou a j -tou detekciou:

$$d^{(2)}(i, j) = \min\{1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_k\} \quad (4.3)$$

Pre porovnávanie vzhľadového modelu je navrhnutá rovnaká rozhodovacia logika, avšak $t^{(2)}$ hranica, musí byť nájdená pomocou trénovania neurónovej siete na vhodných dátových sadách.

Kapitola 5

Rozpoznávanie činností

Výsledkom tejto práce je práve rozpoznávanie činností (sekcia 5.1), ktoré je zabezpečené pomocou dvoch spôsobov. Prvým spôsobom je rozpoznávanie jednoduchých akcií neurónovou sietou, ktorá je objasnená v sekcií 5.4. Pre správne pochopenie fungovania tohto modelu je nutné sa najprv zoznámiť s problematikou spojenou práve s prístupom neurónových sietí k rozpoznávaniu akcií. O tejto problematike a výzvach pojednávajú sekcie 5.2 a 5.3. Druhým použitým spôsobom je spôsob spájania sémantických informácií získavaných z predchádzajúcej analýzy videa modelmi pre rozpoznávanie a sledovanie objektov. Tento spôsob je ďalej preberaný v sekcií 5.5.

5.1 Úvod do rozpoznávania akcií

Rozpoznávanie akcií je veľmi dôležitou disciplínou v dnešnom svete. Napríklad v prípadoch, kedy chodci prechádzajú cez prechod pre chodcov, je potrebné, aby bol vodič motorového vozidla schopný vyhodnotiť, že je človek otočený k ceste a že jeho úmyslom je prejsť cez ňu. Ľudský nervový systém dokáže takéto typy akcií rozpoznať bez väčšieho premýšľania na základe dlhoročných skúseností. Avšak človek môže podliehať rôznym vplyvom ako je únava, stres alebo strata pozornosti. Pokiaľ je človek unavený a zaspáva za volantom, môže sa stať, že nestihne dostatočne včas zareagovať na prechádzajúceho chodca. V najhoršom prípade môže takáto situácia skončiť dokonca smrťou oboch ľudí.

Práve kvôli takýmto situáciám je jedným z veľkých cieľov počítačového videnia vytvoriť systém podobný tomu ľudskému, ktorý by bol schopný úspešne rozpoznávať rôzne situácie vysokou rýchlosťou. Systém schopný rozpoznávať akcie a činnosti ľudí, by dokázal obrovským spôsobom nie len uľahčiť ľudom život, ale aj zabrániť rôznym nechceným situáciám.

V spomínanom príklade s prechodom človeka cez cestu dokáže auto zastaviť na základe informácií z rôznych senzorov merajúcich vzdialenosť od objektov. Avšak v iných situáciach ako sú krádeže alebo vykonávaná trestná činnosť, senzory merajúce vzdialenosť nepomôžu. Preto je potrebné dokázať analyzovať situáciu vyskytujúcu sa vo videu pomocou rôznych techník.

Pre rozpoznávanie a analýzu činností človeka vo videu bolo navrhnutých mnoho techník, akými sú napríklad sledovanie trajektórie ľudí, rozpoznávanie akcií na základe pohybu klíbov človeka alebo rozpoznávanie akcií pomocou extrakcie vizuálneho modelu človeka [27].

Táto disciplína, ktorej podstatou je analyzovať činnosti a situácie človeka, by sa zjednodušene dala rozdeliť do dvoch kategórií:

- rozpoznávanie akcií – rozpoznanie akcie z videa alebo sekvencie obrázkov, na ktorých akcia už bola vykonaná,
- predpovedanie akcií – predpovedanie, že akcia nastane z časovo neúplných video dát.



Obr. 5.1: Znázornenie rozdielu medzi rozpoznávaním akcie a predpovedaním akcie

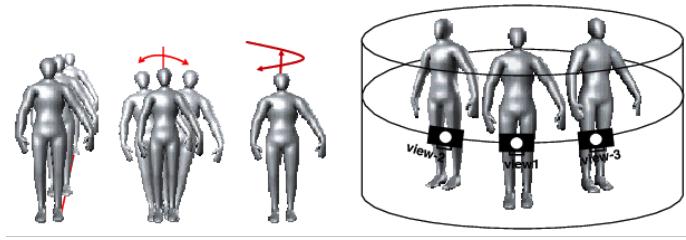
Rozpoznávanie akcií je úlohou počítačového videnia, kedy je akcia rozpoznaná z aktuálneho stavu diania vo videu. Inak povedané, akcia je rozpoznaná až zo získaných informácií po tom, čo jej vykonávanie bolo dokončené. V niektorých prípadoch ako je krádež alebo zrážka dvoch áut, môže byť už neskoro pre rozpoznávanie takýchto situácií. V tom prípade je vhodnejšou alternatívou predpovedanie akcií, ktorej úlohou je predpovedať danú situáciu v dostatočnom predstihu [27] (obrázok 5.1).

5.2 Výzvy metód rozpoznávajúcich akcie

Na vzdory tomu, že algoritmy rozpoznávajúce akcie pokročili miličnými krokmi, stále existujú rôzne ovplyvňujúce faktory, ktoré zatiaľ nie je možné úplne eliminovať. Akcie vykonávané človekom sa líšia v rôznych formách. Problémami môžu byť rôzne pozície človeka, jeho rýchlosť alebo uhlosť, pod ktorým je daná akcia snímaná (obrázok 5.2). Preto je aj rozpoznanie jednoduchých akcií, akými sú napríklad behanie a kráčanie veľmi ľahko rozlíšiteľné, keďže pri nich človek vykonáva pomerne rovnaký pohyb, no s rozličnými rýchlosťami [30].

Rozpoznávanie akcií býva často úzko späté s rozpoznávaním objektov a ich sledovaním. Pokial má model o nejakom objekte vedieť povedať, akú akciu vykonával, je potrebné, aby bol schopný rozpoznať jeho pozíciu po celú dobu vykonávanej akcie. Preto ďalšími problémami, ktorým rozpoznávanie objektov podlieha, sú problémy pri detekcii objektov, akými sú napríklad zmeny osvetlenia objektu alebo prekrytie objektu [70].

Velkým problémom je aj výber jednotlivých obrázkov. Akcia alebo činnosť bývajú rozpoznávané zo sekvencie nadväzujúcich obrázkov videa. Avšak nie všetky obrázky majú vhodný popisný charakter pre rozpoznanie akcie. Je zbytočné používať obrázky, na ktorých sa scéna skoro vôbec nezmenila, a preto je potrebné preskakovať určitý počet obrázkov.



Obr. 5.2: Znázornenie rôznych pôz a uhlov pohľadu pri vykonávaní akcie „kráčanie“¹

Naviac sa jednotlivé akcie líšia náročnosťou rozpoznania. Niektoré je možné rozpoznať už z jedného obrázka, iné potrebujú dlhšiu sekvenciu [24].

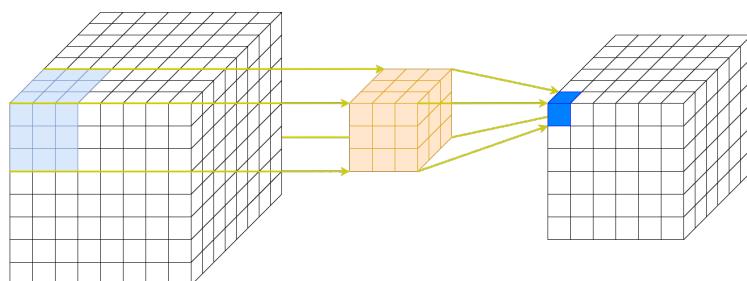
Tieto problémy výrazne stážujú rozpoznávanie akcií, a preto je táto disciplína veľmi náročná. Vytvoriť všeobecný systém, ktorý by bol schopný rozpoznávať všetky alebo aspoň viaceré druhy akcií, je takmer nemožné.

5.3 Prístup neurónových sieti k rozpoznávaniu akcií

Pred neurónovými sietami bolo používaných mnoho rôznych techník, ktoré sa spoliehali na ručne modelované časovo-priestorové reprezentácie akcií. Tento prístup však vyžadoval expertnú znalosť vo vytváraní metód pre ich extrakciu. Zároveň tieto algoritmy neboli dostatočne všeobecné, aby dokázali rozpoznávať rôzne variácie akcií a boli odolné voči rôznym vplyvom.

V poslednej dobe sa však vysokému záujmu tešia práve prístupy založené na hlbokom učení, práve pre svoju schopnosť samostatne modelovať všeobecné reprezentácie akcií, bez nutnosti nastavovať rôzne parametre. Dvomi hlavnými premennými pri vytváraní týchto neurónových sietí sú konvolučná operácia a modelovanie časovej informácie.

Konvolučná operácia, ako bolo spomenuté v sekcií 3.2, je operácia, ktorá zoskupuje hodnoty pixelov z malých susedných oblastí obrázka, na základe ktorých modeluje výslednú reprezentáciu obrázka pomocou konvolučného filtra. Algoritmy pre rozpoznávanie akcií však vyžadujú aj modelovanie časovo-priestorovej reprezentácie, a preto niektoré techniky nahradzajú 2D konvolúciu zložitejšou 3D konvolúciou (obrázok 5.3).



Obr. 5.3: Zjednodušený pohľad na operáciu 3D konvolúcia

¹prevzaté z https://www.researchgate.net/figure/Variabilities-of-human-actions-and-Anthropometry-variation-images-with-different-body_fig1_221292659

3D konvolúcia je operácia podobná 2D konvolúcií, avšak je vykonávaná na viacerých obrázkoch naraz. Takáto operácia je schopná modelovať hierarchické časovo-priestorové reprezentácie, ale obsahuje omnoho viacej parametrov, čo výrazne stáže proces učenia oproti 2D konvolučným neurónovým sietiam.

Ďalšou klúčovou premennou pri zostavovaní modelov pre rozpoznávanie akcií je časové modelovanie. Pre toto modelovanie sú najčastejšie používané tri druhy sietí:

- časovo-priestorové siete,
- viac-prúdové siete,
- hybridné siete.

Každá z týchto sietí sa vyznačuje nejakými znakmi, výhodami alebo nevýhodami. Časovo-priestorové siete sú považované za rozšírenie 2D konvolučných neurónových sietí a pre modelovanie časovej reprezentácie využívajú vyššie spomínanú 3D konvolúciu. Viac-prúdové siete, ako z názvu vyplýva, používajú viacero prúdov, ktorými prechádzajú jednotlivé obrázky zo sekvencie. Tieto prúdy sú konvolučné siete. Jedna sieť je využívaná pre získanie informácie o statickej časti obrázku, čo znamená, že zistuje, ktoré časti obrázka sú nehybné a druhá sieť vykonáva rozpoznávanie akcie na základe vykonávaného pohybu. Skupinu uzatvárajú hybridné siete, ktoré pre modelovanie časovej informácie využívajú rôzne druhy sietí. Najčastejšie sa vyskytujú v spojení konvolučná neurónová sieť a LSTM rekurventná neurónová sieť, ktorá je schopná naučiť sa postupné závislosti v sekvenciách dát. Viac o prístupoch neurónových sietí pre rozpoznávanie akcií je možné sa dozvedieť v článku *Going deeper into action recognition* [22].

5.4 Model SlowFast

Pre analýzu videa bolo navrhnutých mnoho techník a modelov neurónových sietí. Avšak pokiaľ je potrebné rozpoznávať akcie z určitých oblastí, ako sú napríklad ohraničujúce boxy obsahujúce osoby, je vhodných len niekoľko prístupov. V tejto práci je použitý model *SlowFast* pre svoje vynikajúce výsledky v oboch oblastiach, či už klasifikácie akcií alebo rozpoznávania akcií v oblastiach z videa.

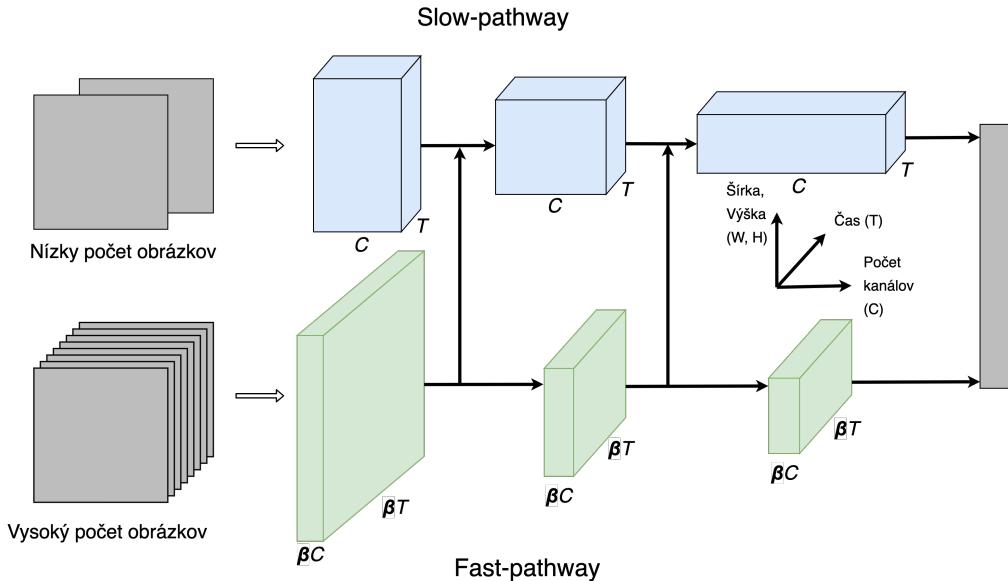
Model sa zameriava na myšlienku, že jednotlivé informácie sa vo videu menia rôznymi rýchlosťami. Napríklad, pri človeku, ktorý máva rukami, je jasné, že sa typ objektu nezmení a jeho sémantické informácie ako farba, nasvietenie alebo textúra sa nemenia s vysokou rýchlosťou. Preto tieto informácie stačí obnovovať pomaly. Naopak vykonávaný pohyb sa vyvíja omnoho rýchlejšie a z toho dôvodu je potrebné ho zachytávať častejšie.

Na základe týchto myšlienok model *SlowFast* predstavuje dve cesty pre rozpoznávanie akcií. Prvá cesta nazývaná *Slow-Pathway* slúži pre zachytenie sémantických informácií, ktoré môžu byť získané s jedného obrázka alebo malej sekvencie obrázkov. Druhá cesta nazývaná *Fast-Pathway* je využívaná pre zachytenie rýchlych zmien pohybu. Na rozdiel od *Slow-Pathway* využíva vysokú obnovovaciu rýchlosť a veľké časové rozlíšenie. Tieto cesty sú spájané pomocou neskorších prepojení [72] (obrázok 5.4).

Slow-Pathway. *Slow-Pathway* cesta môže byť predstavená ktoroukoľvek konvolučnou neurónovou sietiou, ktorá dokáže pracovať so sekvenciami videa ako s časovo-priestorovými dátami. Klúčovým aspektom tejto cesty sú veľké kroky τ medzi vstupnými obrázkami, čo znamená, že daná cesta spracováva informácie iba z jedného obrázku z τ obrázkov. Pokial je

označený počet vzoriek obrázkov spracovaných cestou *Slow-Pathway* ako T , výsledný počet obrázkov v sekvencií je $T \times \tau$.

Fast-Pathway. *Fast-Pathway* je predstavená pomocou konvolučnej siete, ktorá musí disponovať určitými vlastnosťami. Cesta využíva menšie časové kroky τ/α medzi obrázkami, kde $\alpha > 1$ je pomer počtu obrázkov medzi *Slow-Pathway* a *Fast-Pathway*. Obe cesty spracúvajú rovnakú sekvenčiu obrázkov, avšak *Fast-Pathway* vyzorkuje αT obrázky, α -krát hustejšie ako *Slow-Pathway*. Ďalšou vlastnosťou tejto siete je, že využíva nielen vysoké vstupné rozlíšenie, no takéto rozlíšenie aj propaguje jednotlivými vrstvami siete. Nepoužíva žiadne časové vrstvy znižujúce kvalitu vzoriek až po dobu klasifikácie. Poslednou vlastnosťou, ktorú táto sieť musí splňať, je znížený počet kapacity kanálov jednotlivých konvolučných filtrov. Pomer kapacity kanálov β medzi *Fast-Pathway* a *Slow-Pathway* musí dodržiavať podmienku $\beta < 1$. Typicky je tento pomer nastavený ako $\beta = 1/8$. Takéto zníženie počtu kapacity kanálov zachová časové rozlíšenie, čo umožňuje tejto konvolučnej sieti detailnejšie zachytiť pohyb.



Obr. 5.4: Ukážka ciest modelu *SlowFast* a ich neskorších prepojení

Neskoršie prepojenia. Medzi sietami je potrebné zabezpečiť presun informácií a reprezentácií naučených druhou cestou. Táto schopnosť je implementovaná pomocou neskorších prepojení, ktoré slúžia pre spájanie dvojprúdových sietí založených na optickom toku. Model *SlowFast* využíva jednosmerné prepojenie, ktoré pripája informácie získané z *Fast-Pathway* do *Slow-Pathway* cesty, ako je znázornené na obrázku 5.4.

Výstup obidvoch ciest je spojený do výsledného vektoru reprezentujúceho akciu, pomocou vrstvy *global average-pooling*. Po spojení je tento vektor vložený do poslednej plne prepojenej vrstvy, ktorá je zodpovedná za výslednú klasifikáciu akcie zo sekvencie obrázkov [16, 66].

5.5 Rozpoznávanie akcií pomocou kompozície sémantických informácií

Jedným z ďalších spôsobov, ktorý je použitý v tejto práci a akým je možné rozpoznávať akcie a analyzovať správanie ľudí, je spájanie informácií. Tento proces rozpoznávania akcií sa pohráva s myšlienkom, že sa pri jednotlivých akciách vykonávaných človekom opakujú informácie, ktoré je možné získať pomocou jednoduchej analýzy videa. Takéto získané informácie je možné skladať do komplexnejších a zložitejších reprezentácií činností, aké nie sú dnešné modely neurónových sietí schopné rozpoznávať.

Príkladom zloženej činnosti môže byť napríklad vystupovanie človeka z auta. Pre rozpoznanie tejto akcie je potrebné si uvedomiť, čo sa v danej situácii deje. Modely pre rozpoznanie a sledovanie objektov prvýkrát zaznamenajú človeka až po vystúpení z auta, čo znamená, že jeho trajektória začína v jeho okolí. Spolu s informáciou o smere, v ktorom sa človek nasledujúce detekcie vydáva, je možné takúto situáciu vyhodnotiť ako vystupovanie z auta.

V predchádzajúcich kapitolách 3 a 4 boli vysvetlené modely, pomocou ktorých je možné získavať práve takéto informácie, akými sú pozícia objektu, jeho typ a iné. Najhlavnejšou informáciou používanou v tejto práci, ktorú je možné odvodíť, je trajektória objektov. Práve trajektória je klíčovou informáciou pri definovaní jednotlivých činností, pretože je pomocou nej možné logicky odvodiť rôzne informácie, ako sú napríklad smer alebo rýchlosť pohybu objektu.

Trajektória objektu je zložená z bodov v určitých časových úsekokoch. Vďaka časovej informácii môžeme nad bodmi v určitých časoch vykonávať rôzne operácie, akými sú porovnávanie vzdialenosť alebo uhlovej veľkosti medzi smermi objektov. Na základe týchto jednoduchých operácií, potom systém odvoduje výskyt definovanej činnosti, pokiaľ sú splnené určité podmienky medzi objektmi v čase.

V tejto práci sú využívané konkrétnie štyri operácie, pomocou ktorých sú spracovávané body trajektórie a ktoré napomáhajú pri určovaní a rozhodovaní o výskytu situácie vo videu:

- vzdialenosť medzi objektmi,
- porovnávanie uhla medzi objektmi,
- výskyt objektu v polygóne,
- porovnávanie trajektórií.

Pre výpočet vzdialenosťi medzi objektmi je používaný jednoduchý vzorec:

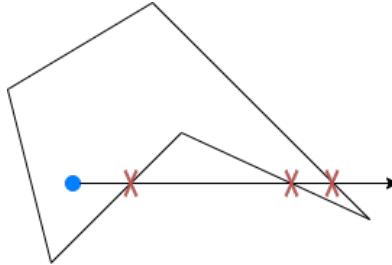
$$\text{Vzdialenosť}_{(p_1, p_2)} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5.1)$$

kde x_1, y_1 označujú p_1 pozíciu prvého a x_2, y_2 označujú p_2 pozíciu druhého stredného bodu objektu na obrázku, získaného z ohraňujúceho boxu [11].

Pomocou porovnania uhla medzi objektmi je možné zistiť, či sa človek pohybuje smerom k inému objektu. Pri porovnaní je potrebné definovať dve čiary, medzi ktorými je vypočítaný uhol. Prvú čiaru, $l_1 = [[x_1, y_1], [x_2, y_2]]$, tvoria dva body, stredový bod objektu človeka a bod vypočítaný pomocou smeru a rýchlosťi jeho pohybu. Druhá čiara $l_2 = [[x_3, y_3], [x_4, y_4]]$ je tvorená stredovým bodom objektu človeka a stredovým bodom objektu, ku ktorému chceme zistiť výsledný uhol. Výsledný uhol medzi čiarami je vypočítaný pomocou vzorca:

$$\text{Uhol}_{(l_1, l_2)} = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \quad (5.2)$$

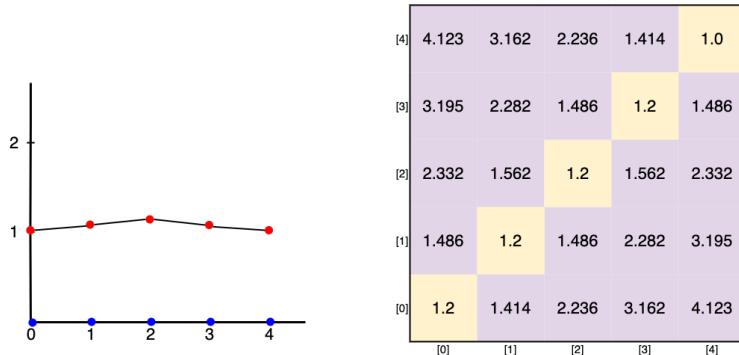
kde $\mathbf{a} = [x_2 - x_1, y_2 - y_1]$, $\mathbf{b} = [x_4 - x_3, y_4 - y_3]$, $\mathbf{a} \cdot \mathbf{b}$ je skalárny súčin medzi vektormi vypočítaný ako $\mathbf{a} \cdot \mathbf{b} = a_x \times b_x + a_y \times b_y$ a $|\mathbf{a}| |\mathbf{b}|$ je súčin magnitúd vektorov vypočítaný pomocou vzorca $|\mathbf{a}| = \sqrt{a_x^2 + a_y^2}$. Výsledná hodnota uhla je vyjadrená pomocou stupňov [52].



Obr. 5.5: Znázornenie algoritmu pre odhalenie bodu v polygóne

Výskyt objektu v polygóne je kontrolovaný pomocou algoritmu, ktorého vstupom je polygóon a stredový bod objektu. Algoritmus postupne zvyšuje hodnotu osy x od daného bodu a počíta kolkokrát nastane prenutie s hranou polygónu (obrázok 5.5). Nepárny počet prenutí značí, že sa daný bod nachádza v polygóne [19].

Poslednou z operácií používanou v tejto práci je porovnávanie trajektórií. Vybraným algoritmom bol práve algoritmus s názvom *vzdialenosť Fréchet*, ktorý dokáže porovnávať trajektórie rôznych dĺžok.



Obr. 5.6: Ukážka grafu dvoch trajektórií a prechodu *free-space* diagramom pre vypočítané vzdialenosť medzi bodmi trajektórií

Výpočet *vzdialenosťi Fréchet* je uskutočnený pomocným poľom, ktoré sa nazýva *free-space diagram* (obrázok 5.6). Toto pole je štvorcového tvaru a ukladá Euklidovské vzdialenosťi medzi bodmi oboch trajektórií. Pole obsahuje všetky možné kombinácie bodov, medzi ktorými sú vypočítané vzdialenosťi, pre ktoré platia dve pravidlá. Nie je možné vracať sa späť v trase a počiatočné a koncové body sú vždy zhodné. Veľkosť takého poľa je maximálne $p \times q$ kde p a q označujú dĺžku trajektórií [15].

Pre výpočet vzdialenosťi *Fréchet* je nutné prejsť celé pole s uloženými Euklidovskými vzdialenosťami medzi bodmi, kedy prechod poľa začína v lavom dolnom rohu poľa označenom indexmi ($i = 0, j = 0$). Pri postupe poľom sú brané tri hodnoty vzdialenosťí

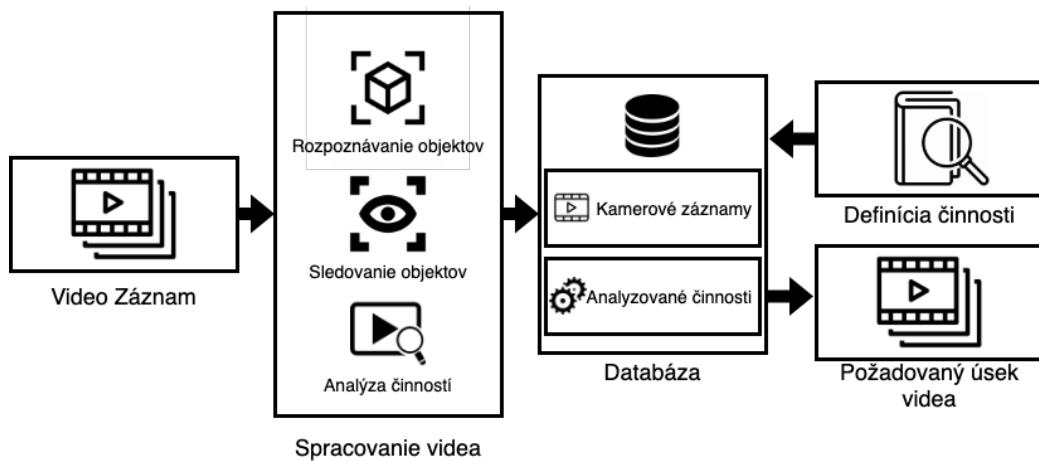
na pozíciach $(i + 1, j)$, $(i + 1, j + 1)$ a $(i, j + 1)$, z ktorých je vyberaná minimálna hodnota. Výpočet končí pokiaľ sa algoritmus dostane na prvok v pravom hornom rohu. Výsledná vzdialenosť medzi trajektóriami je maximálna vzdialenosť z hodnôt, na ktorých sa algoritmus nachádzal pri prechode poľom [3].

Kapitola 6

Návrh a implementácia

Práca sa zameriava na vytvorenie časti systému pre prehľadávanie videí, zodpovednej za získavanie informácií o objektoch pomocou analýzy videa neurónovými sietami. Informácie sú po získaní uložené do databázy. Následne sú pomocou vzťahov medzi informáciami definované typy činností človeka. Vzťahy medzi informáciami sú vyhľadávané v databáze, na základe čoho, sú potom získané úseky videa s požadovanou činnosťou. Táto kapitola sa zameriava na návrh a implementáciu jednotlivých častí systému zodpovedných za analýzu videa, definíciu niektorých činností človeka a získanie úsekov videa s definovanou činnosťou.

Pre rozpoznávanie komplexných akcií je veľmi dôležité uvedomiť si, akým spôsobom by mal tento proces prebiehať. Niektoré práce ako napríklad *Hierarchical recurrent neural network for skeleton based action recognition* [14] sa zameriavajú na rozpoznanie akcií z jednoduchej kostry človeka a jej pozície. V tejto práci sa však pre zjednodušenie a zrýchlenie procesu využíva rozpoznávanie akcií na základe vyhodnotenia vzťahov medzi získanými sémantickými informáciami. Zároveň však tento spôsob umožňuje modelovať viacero činností, ktoré nie sú schopné rozpoznať ani novodobé neurónové siete. O niektorých ďalších metódach rozpoznávania akcií je možné sa dozvedieť v článku *Going deeper into action recognition* [22].



Obr. 6.1: Schéma systému znázorňujúca postupné prúdenie dát jednotlivými modulmi

Dôležitou vlastnosťou systému pre prehľadávanie videa je schopnosť rozpoznať jednotlivé sémantické informácie, akými sú napríklad pozícia objektu v rámci obrázka, trieda daného objektu alebo jeho veľkosť. Pomocou týchto informácií je možné ďalej odvodíť niektoré

podstatné časovo priestorové informácie, akými sú smer pohybu a trajektória objektu počas výskytu na videu.

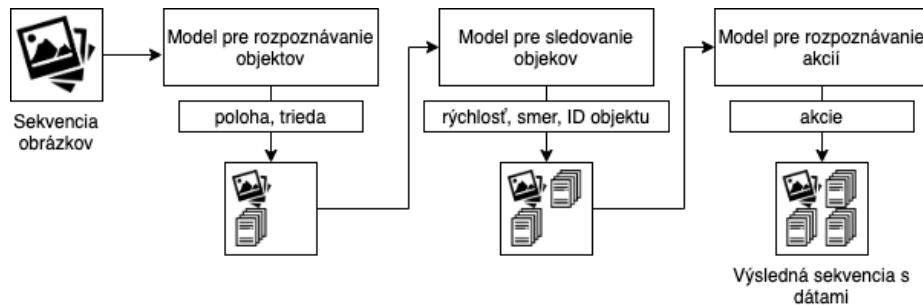
Systém bude v nasledujúcich sekciách popísaný v smere prúdenia dát jednotlivými časťami systému. Jednoduchá schéma výsledného systému, ktorú je možné vidieť na obrázku 6.1, znázorňuje toto prúdenie dát. Dáta v podobe videozáznamu sa postupne ako sekvencie obrázkov dostávajú do modulu pre spracovanie videa, ktorý je zodpovedný za získanie jednotlivých informácií o objektoch vyskytujúcich sa vo videu. Získané informácie sa potom ukladajú do databázy. Po uložení vyhľadávajú definované činnosti v databáze informácie, na základe ktorých určujú, či sa v danom videozázname činnosť nachádza. Výsledkom je úsek videa s požadovanou činnosťou.

6.1 Spracovanie dát

Systém využíva dátá, ktoré musia byť pred vstupom do modelov upravené. Pre systém sú vstupnými dátami videozáznamy rôzneho rozlíšenia, dĺžky alebo vzorkovacej rýchlosťi obrázkov za sekundu. Dáta sú postupne spracovávané v závislosti na použití. Pre spracovávanie video-dát bola použitá voľne dostupná knižnica *OpenCV*, slúžiaca pre úpravu vizuálnych dát. Pomocou tejto knižnice boli jednotlivé videá rozdelené na menšie sekvencie obrázkov podľa nastavenej dĺžky.

Každá sekvencia obrázkov je predstavená pomocou objektu. Objekt drží informácie o začiatokom indexe sekvencie obrázkov v rámci videa a zároveň obsahuje všetky obrázky danej sekvencie. Postupne je tento objekt pri každom prechode niektorým z modelov obohatený o informácie získané práve z použitého modelu, ako sú napríklad jednotlivé triedy, ohraňujúce boxy objektov alebo identifikačné čísla objektov.

Po každom vytvorení sekvencie obrázkov, sú nad touto sekvenciou postupne spúštané jednotlivé modely. Na obrázku 6.2 je možné vidieť ako objekt postupuje modelmi, pričom získava dátá.



Obr. 6.2: Ukážka prúdenia dát a sekvencie obrázkov jednotlivými modelmi systému

Výsledné získané dátá sú vo forme *Python* slovníku. Tento slovník je súčasťou objektu reprezentujúceho sekvenciu obrázkov a je po každom prechode niektorým z modelov obohatený o dátá získané z daného modelu. Následne sa po ich získaní dátá zo slovníka spracujú do podoby, ktorá im umožňuje, aby boli uložené do súboru.

6.2 Model pre rozpoznávanie objektov

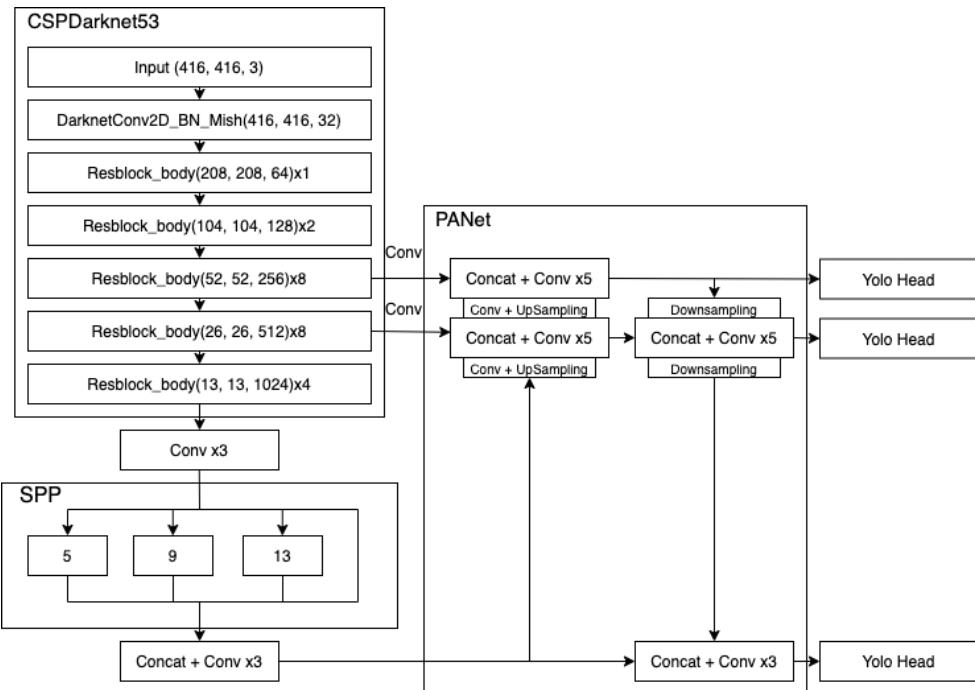
Prvou časťou systému, ktorá spracováva video, je model pre rozpoznávanie objektov. Jeho výber ovplyvňuje presnosť a schopnosť výsledného systému rozpoznávať komplexné činnosti

ľudí. Kedže sú vstupom pre model videá rôznych veľkostí, dĺžok a formátov, je potrebné tomu prispôsobiť aj jeho výber.

Na základe vstupných dát je nutné, aby bol vybraný model schopný rozpoznávať jednotlivé objekty zo sekvencie obrázkov videa s vysokou rýchlosťou a zároveň dostatočnou presnosťou. Preto bol pre túto prácu vybraný model neurónovej siete s názvom *You Only Look Once*.

You Only Look Once (ďalej len *YOLO*) je model využívajúci jednu fázu pre rozpoznávanie objektov. Konkrétnie ide o štvrtú verziu tohto modelu. V oficiálnej práci *YOLOv4: Optimal Speed and Accuracy of Object Detection* [8] autor popisuje jednotlivé časti, z ktorých je zložený. Za zmienku určite stojí jeho schopnosť rozpoznávať objekty s vysokou rýchlosťou a presnosťou, porovnatelnou s ostatnými dvojfázovými modelmi.

Model *YOLO* štvrtej verzie (obrázok 6.3) v porovnaní so svojím predchodom prešiel rôznymi zmenami. Skladá sa z troch častí, ktoré sa nazývajú chrbtica, krk a hlava modelu (v angl. backbone, neck a head). Vstupný obrázok najprv putuje jednotlivými vrstvami chrbtice, ktorá je implementovaná pomocou *CSPDarknet-53* neurónovej siete. Získané mapy popisných vlastností sú vstupom pre krk siete, ktorý je predstavený pomocou *Spatial Pyramid Pooling* vrstvy, zodpovednej za zvýraznenie najvýznamnejších znakov vlastností. Nakoniec namiesto *Feature Pyramid siete* použitej v tretej verzií *YOLO* modelu, nástupná verzia používa *Path Aggregation siet* [38].



Obr. 6.3: Architektúra modelu You Only Look Once verzie 4

V oficiálnych článkoch k modelom pre rozpoznávanie objektov ako napríklad *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks* [48], autori často porovnávajú svoj model práve s modelmi *YOLO* alebo modelmi založenými na regiónových konvolučných neurónových sieťach (v angl. *Region Based Convolution Neural Networks*). V porovnaní s najnovšou verzou modelu regiónových konvolučných neurónových sietí, dokáže *YOLO* rozpoznávať objekty s niekoľko násobnou rýchlosťou. Na dátovej

sade *MS COCO* [61] pomocou vyhodnocovacej techniky *Mean Average Precision*, dokázal *YOLO* model štvrej verzie dosiahnuť presnosť 65.7% pri rýchlosťi 33 obrázkov za sekundu. Pre porovnanie, model *Faster R-CNN* dosiahol presnosť 76.1% na rovnakej dátovej sade, no jeho maximálna rýchlosť dosahovala hodnoty len 7 obrázkov za sekundu.

Pre vybraný model existuje mnoho vytvorených voľne dostupných implementácií, a preto ho netreba na novo programovať. Zároveň, keďže je úlohou tejto práce rozpoznávanie činností a nie implementácia modelu pre rozpoznávanie objektov, bola implementácia tohto modelu prevzatá z repozitáru stránky *Github*¹. Model siete *YOLO* je implementovaný pomocou voľne dostupnej knižnice strojového učenia *Pytorch* a programovacieho jazyka *Python*.

Rozpoznávanie objektov vybraným modelom prebieha postupne na každej sekvencii obrázkov videa. Pre zrýchlenie implementácie je navrhnutý systém preskakovania obrázkov. Tento systém určuje počet obrázkov zo sekvencie, ktoré majú byť preskočené, aby boli informácie o polohe objektu počas jednej sekundy videa stále aktuálne.

Pre použitie modelu je potrebné upraviť jednotlivé vstupné obrázky. Tieto obrázky sú najprv zmenšené na určitú veľkosť, ktorú vyžaduje model neurónovej siete. Následne nad obrázkami prebehne konverzia z farebného systému BGR (modrá, zelená, červená), na systém RGB (červená, zelená, modrá). Po vykonaní rozpoznania jednotlivých objektov, sa získané informácie ukladajú do dátovej štruktúry v podobe *Python* slovníka, uchovávajúcej polohu, triedu a pravdepodobnosť výskytu každého objektu. Tento slovník je vložený do objektu reprezentujúceho sekvenciu obrázkov.

6.3 Model pre sledovanie objektov

Ak má byť výsledný systém schopný vyhodnotiť jednotlivé činnosti ľudí, je potrebné, aby dokázal určiť, ktorý človek danú akciu vykonával. Z toho dôvodu musí vedieť sledovať jednotlivé objekty a priradiť im určitú formu unikátneho identifikátoru. V tomto prípade výsledný systém ako formu identifikátoru používa identifikačné číslo, ktoré je priradené práve jednému objektu. O priradenie tohto čísla sa stará sledovací algoritmus.

V dnešnej dobe existuje viaceré spôsobov, akým sa dajú jednotlivé objekty v rámci videa sledovať. Počas posledných rokov bolo navrhnutých mnoho rôznych algoritmov pre sledovanie objektov. Niektoré algoritmy ako napríklad *Tracktor* [6] alebo *Track-RCNN* [56] sú priamo spojené s modelmi pre rozpoznávanie objektov. Iné, *SiamMask* [63], nedokážu sledovať viaceré objekty s vyhovujúcou rýchlosťou. V tejto práci bol použitý pre sledovanie objektov algoritmus *Deep Sort*, ktorý zabezpečuje dostatočnú mieru presnosti v pomere s rýchlosťou sledovania objektov v skutočnom čase. Dôležitá je práve rýchlosť, a to z toho dôvodu, že výsledný systém musí byť schopný spracovávať videá veľkých dĺžok, akými videozáznamy disponujú.

Algoritmus *Deep Sort* je už druhou verziou tohto sledovacieho algoritmu. V prvej verzii s názvom *Sort*, algoritmus používal pre sledovanie objektov ako vstup ohraničujúci box, z ktorého dokázal pomocou *Hungarian algoritmu* priradiť box k objektom z predchádzajúceho snímku. Využitím *Kalmanovo filtra* predpovedal zmenu pohybu daného objektu [7]. Kvôli problému s pomerne častými výmenami jednotlivých identifikátorov objektov, novšia verzia algoritmu s názvom *Deep Sort* využíva neurónovú sieť, ktorá popisuje vzhľad objektov. V kombinácii s informáciami o pohybe je model schopný obmedziť problém so zámenami jednotlivých identifikačných čísel objektov [65].

¹<https://github.com/Tianxiaomo/pytorch-YOLOv4>

Porovnania výsledkov testovania sledovacích algoritmov na dátovej sade *MOT16* [2], ukazujú, že *Deep Sort* dosahuje podobné hodnoty ako ostatné porovnávané algoritmy, a že novšia verzia algoritmu dokázala úspešne obmedziť počet zmien identifikátorov z pôvodných 1423 na 781. O ďalších výsledkoch testovania jednotlivých sledovacích algoritmov je možné sa dočítať v článku *Simple online and realtime tracking with a deep association metric* [65].

Pri programovaní výsledného systému bola implementácia algoritmu *Deep Sort* získaaná z repozitáru stránky *Github*². Vstupom pre tento algoritmus sú ohraničujúce boxy, získaané pomocou modelu pre rozpoznávanie objektov. Táto implementácia uchováva informácie o každej jednej trase vykonávanej rozpoznanými objektami. Pomocou uchovávaných informácií sú ďalej získavané:

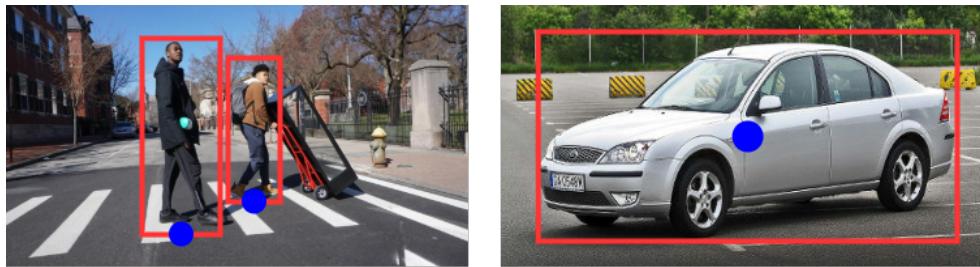
- stredný bod objektu,
- smer pohybu,
- rýchlosť pohybu.

Stredný bod objektu je získaaný z ohraničujúceho boxu každého objektu v čase, kedy je objekt sledovaný. Tento stredný bod slúži pre odvodenie smeru, rýchlosťi pohybu a trajektorie, ktorá je získavaná pomocou databázy. Stredný bod sa líši na základe typu objektu. Pokiaľ ide o typ objektu človek, jeho stredný bod v čase t je vypočítaný pomocou vzorca:

$$[x, y](t) = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x_1+x_2}{2} \\ y_2 \end{bmatrix} \quad (6.1)$$

kde x_1, x_2 označujú počiatočný a koncový bod ohraničujúceho boxu horizontálnej súradnice a y_1, y_2 označujú počiatočný a koncový bod ohraničujúceho boxu vertikálnej súradnice (obrázok 6.4). Pokiaľ však ide o objekt iného typu ako človek, jeho stredný bod je vypočítaný úpravou vzorca:

$$[x, y](t) = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x_1+x_2}{2} \\ \frac{y_1+y_2}{2} \end{bmatrix} \quad (6.2)$$



Obr. 6.4: Ukážka stredných bodov podľa typu objektu

Smeru pohybu je odvodený na základe posledných získaných informácií. Pre jeho získaanie je potrebné najprv vypočítať aspoň dva stredné body objektu, a to jeden v čase t , označujúcim aktuálne prebiehajúci čas sledovania objektu a druhý v čase $t - 1$, ktorý označuje stredný bod objektu v minulom čase. Medzi týmito bodmi je vypočítaný uhol. V prvom cykle, kedy je objekt rozpoznaný, nie je možné smer pohybu odvodiť, pretože nie je známa jeho predchádzajúca poloha.

²https://github.com/nwojke/deep_sort

Získané ohraničujúce boxy nie vždy presne odpovedajú ohraničeniu daného rozpoznaného objektu, a to z dôvodu malých nepresností modelu pre rozpoznávanie objektov. Preto je ukladaných niekoľko posledných stredných bodov, ktoré vytvárajú krátku trajektóriu objektu. Táto trajektória je rozdelená na dve polovice, z ktorých je vypočítaný aritmetický priemer z vertikálnej zložky bodov x a horizontálnej zložky bodov y . Následne je vypočítaný uhol medzi dvoma polovicami bodov trajektórie, čím je zabezpečené vyhľadenie výsledného smerového vektoru a jeho hodnôt v rôznych časoch.

Rýchlosť pohybu je určovaná v počte prejdených pixelov od poslednej detekcie. Je vypočítaná určením vzdialenosť medzi dvomi stredovými bodmi v časoch t a $t - 1$. Zároveň ako aj pre smer pohybu, tak aj pre rýchlosť objektu je využívaná korekcia medzi viacerými bodmi uchovávanej krátkej trajektórie objektu.

Všetky výstupné dáta získané z modelu (identifikačné číslo, smer pohybu, rýchlosť pohybu) sú ukladané do dátovej štruktúry so získanými informáciami z modelu pre rozpoznávanie objektov.

6.4 Model pre rozpoznávanie akcií

Disciplína rozpoznávania ľudských akcií je v dnešnej dobe veľmi rozšírenou. Počas posledných rokov bolo predstavených mnoho spôsobov, akými môžu byť ľudské akcie rozpoznané. Napríklad metóda navrhnutá v článku *Joint Distance Maps Based Action Recognition With Convolutional Neural Networks* [31] sa zameriava na detekciu jednoduchej kostry človeka, z ktorej extrahuje pohyb jednotlivých klíbov, na základe čoho pomocou neurónovej siete klasifikuje výslednú akciu. Avšak s navrhnutým modelom pre rozpoznávanie objektov je výhodnejším spôsobom využiť neurónovú sieť, ktorej vstupom sú anotačné dáta získané z tohto modelu. Viacej o metódach rozpoznávania akcií pojednáva článok *Going deeper into action recognition* [22].

Rozpoznávanie akcií je v tomto systéme navrhnuté dvomi spôsobmi. Prvý spôsob je založený na rozpoznávaní jednoduchých akcií pomocou modelu neurónovej siete a je používaný najmä pre doplnenie informácií. Druhý spôsob je založený na definovaní a vyhodnocovaní vzťahov medzi získanými informáciami o objektoch. Návrh a implementácia druhého spôsobu bude neskôr vysvetlená v sekcii 6.6.

Systém používa pre rozpoznanie niektorých jednoduchých akcií neurónovú sieť s názvom *SlowFast*. Vo svojom oficiálnom článku *SlowFast Networks for Video Recognition* [16], autori porovnávajú efektivitu tohto modelu na dátových sadách *Kinetics 400* [26] a *AVA Dataset* [17], na ktorých dosahujú vynikajúce výsledky. Model *SlowFast* dokázal efektivitou prekonať všetky ostatné modely na disciplínach klasifikácie akcie na videu, ale aj detekcie akcie na konkrétnych oblastiach videa.

Vstupom pre túto neurónovú sieť sú ohraničujúce boxy získané z modelu navrhnutého v sekcii 6.2 a sekvencie niekolkých postupne získavaných obrázkov zo vstupného videa. Tieto informácie sú vkladané do dvoch neurónových sietí, ktoré model *SlowFast* používa. Prvá sieť s názvom *FastPathway* slúži pre rozpoznávanie nehybného pozadia, kým druhá sieť s názvom *SlowPathway* rozpoznáva dynamický obsah sekvencií obrázkov a určuje, o akú akciu ide [72].

Implementácia modelu *SlowFast* bola prevzatá z repozitáru stránky Github³. Táto implementácia potrebuje použiť pred klasifikáciou akcie model pre rozpoznávanie objektov

³<https://github.com/facebookresearch/SlowFast>

typu človek, aby dokázala klasifikovať akciu pre každého rozpoznaného človeka. Pôvodná implementácia bola upravená, aby využívala navrhnutý model spomenutý v sekcií 6.2.

6.5 Databáza

Databáza je navrhnutá tak, aby dokázala pracovať s rôznymi geometrickými a geografickými tvarmi, akými sú body, polygóny alebo čiary znázorňujúce trajektóriu objektov. Pre prácu s útvartmi je používané rozšírenie *Postgres* databázy s názvom *Postgis*. Naviac toto rozšírenie ponúka jednotlivé funkcie slúžiace pre vyhodnocovanie vzdialenosťi bodov, uhlovej veľkosti medzi čiarami alebo vzdialenosťi *Fréchet*, ktoré boli vysvetlené v sekcií 5.5.

```
{
    'frame': 8,
    'detections': [
        {
            'id': 9,
            'type': 'person',
            'box': [1022, 778, 1073, 890],
            'detection_score': 0.5438530445098877,
            'direction_angle': 0.0,
            'position_change': 1.0,
            'action': ['sit'],
            'action_score': [0.4870167374610901]
        }
    ]
}
```

Obr. 6.5: Ukážka slovníku pre jeden snímok obrazovky s jedným rozpoznaným objektom uloženým v súbore

Pomocou spracovávania videa modelmi neurónových sietí sú získané informácie o objektoch v rôznych časoch. Tento čas je predstavený pomocou indexu obrázka v rámci spracovávaného videa. Výsledné informácie sú uložené do *Python* slovníku (obrázok 6.5), ktorý je získavaný samostatne pre každú sekvenciu obrázkov. Po doplnení dát z každého modelu sú informácie zo slovníka ukladané do súboru, ktorý obsahuje všetky informácie o každom rozpoznanom objekte vo videu. Následne sú po ukončení analýzy videa informácie zo súboru spracované a ukladané do databázy.

OBJECT_ID	TYPE						
⋮	⋮						
FRAME	OBJECT_ID	OBJECT_SCORE	BOX	DIRECTION_ANGLE	POSITION_CHANGE	ACTION	ACTION_SCORE
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Obr. 6.6: Ukážka tabuľiek, do ktorých sú ukladané dátá zo slovníku

Spracované informácie sú ukladané do tabuľiek znázornených na obrázku 6.6. Tabuľky ukladajú informácie o každom objekte v rozpoznanom čase. O objektoch typu človek sú naviac uložené informácie o vykonávanej jednoduchej akcii, rozpoznanej pomocou modelu pre rozpoznávanie akcií 6.4.

Informácie z tabuľky sú získavané databázovými dotazmi v závislosti na dátach, ktoré sú potrebné pre určenie prítomnosti vzťahu medzi informáciami. Podľa množstva údajov a komplexnosti jednotlivých databázových dotazov (obrázok 6.7) je ovplyvnená aj rýchlosť vyhodnocovania a vyhľadávania činností.

```
WITH object_points AS (
    SELECT t1.object_id, t1.frame, t2.type, t1.direction_angle,
    t1.box, t1.position_change,
    ST_MakePoint((ST_XMax(t1.box) + ST_XMin(t1.box)) / 2,
    CASE WHEN t2.type = 'person' THEN ST_YMax(t1.box)
    ELSE ST_YMax(t1.box) + ST_YMin(t1.box) / 2 END
    ) AS middle_point
    FROM object_time_data_table AS t1
    JOIN object_table AS t2 ON t1.object_id = t2.object_id
),
poly AS (
    SELECT St_GeomFromWKB(...polygon v hecadecimálnom tvare...) AS poly
),
objects_frame_range_in_polygon AS (
    SELECT object_id, int8range(MIN(frame), MAX(frame)) AS range
    FROM object_points, poly
    WHERE ST_Intersects(poly.poly, middle_point)
    GROUP BY object_id
)
SELECT t1.*, t2.type
FROM objects_frame_range_in_polygon AS t1
JOIN object_table AS t2 ON t1.object_id = t2.object_id
WHERE range <> 'empty' ORDER BY t1.object_id;
```

Obr. 6.7: Ukážka komplexnosti databázového dotazu (tučne – Postgres klúčové slová, kurzíva – Postgis funkcie, podčiarknuté – informácie doplnované pomocou programu)

6.6 Definícia činností

Ďalším, hlavným, použitým spôsobom v tejto práci pre rozpoznávanie činností človeka je definícia vzťahu medzi získanými informáciami z analýzy videa. V porovnaní s modelom pre rozpoznávanie akcií je týmto vyjadrením vzťahov medzi jednotlivými informáciami možné definovať komplexnejšie činnosti, v ktorých figurujú viaceré objekty a informácie naraz.

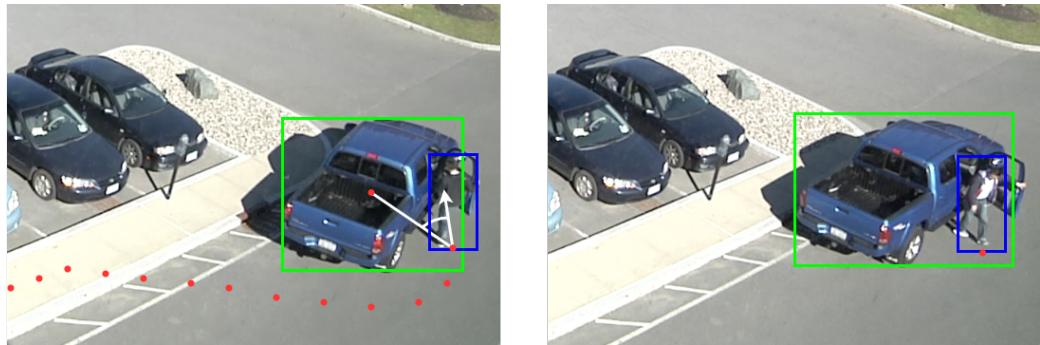
Vzhľadom k miere neistoty ako pri rozpoznávaní jednotlivých informácií, tak pri definícii vzťahov medzi informáciami na udalosti a činnosti, je naviac potreba uvažovať o vyhľadávaní ako o úlohe usporiadania možných výskytov danej udalosti v určitých časoch, podľa miery pravdepodobnosti či istoty, že práve vybraná časť videozáznamu odpovedá hľadanému úseku.

Kedže je možné pomocou vyjadrenia vzťahov medzi získanými informáciami zadefinovať veľké množstvo činností, boli v tejto práci implementované len niektoré, pre získanie obrazu o schopnosti rozpoznať činnosti týmto spôsobom. Zároveň je veľmi náročné nájsť

vhodnú dátovú sadu, ktorá by obsahovala videá s činnosťami, ako napríklad „dobiehanie na električku cez prechod pre chodcov v čase červenej na semafore“. Najväčším problémom sú v takýchto prípadoch právne a etické podmienky, ktoré zakazujú používať dátové sady bez súhlasu ľudí, ktorý sa na daných záznamoch nachádzajú. Preto konkrétnymi činnosťami testovanými v tejto práci sú:

- nastupovanie do auta,
- vystupovanie z auta,
- kráčanie dvoch a viacerých ľudí spolu,
- presúvanie sa medzi definovanými miestami.

Činnosť nastupovania do auta je v tejto práci definovaná spojením informácií o človeku a najbližšom aute k danému človeku. Je možné predpokladať, že pokiaľ trajektória človeka končí v určitej vzdialosti od auta a zároveň je človek otočený smerom k autu, ide práve o spomínané nastupovanie do auta (obrázok 6.8). Miera istoty je pri tejto činnosti vyjadrená aritmetickým priemerom medzi súčtom hodnôt miery istoty rozpoznania objektov človek a auto. Ďalšími faktormi ovplyvňujúcimi mieru istoty je vzdialenosť od auta a smer natočenia k autu.



Nastupovanie do auta

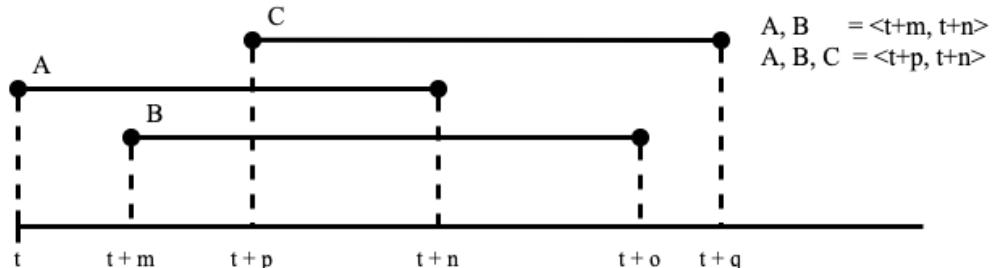
Vystupovanie z auta

Obr. 6.8: Príklad nastupovania a vystupovania z auta, kde červené body označujú trajektóriu objektu a biela šípka označuje porovnanie uhlu medzi bodmi auta, človeka a smerom človeka

Pri vyhľadávaní činnosti vystupovania z auta je možné naopak predpokladať, že trajektória človeka začína v určitej vzdialosti od auta. Kedže pre určenie smeru potrebujeme minimálne dve detekcie, nie je možné porovnávať smer človeka v čase jeho výskytu. Miera istoty je ako aj pri činnosti nastupovania do auta vypočítaná z aritmetického priemeru súčtom hodnôt miery istoty objektov človek a auto. Zároveň pokiaľ stredný bod človeka začína v ohraňujúcim boxe auta, môžeme s najvyššou mierou istoty povedať že z daného auta vystupoval. So zvyšujúcou sa vzdialenosťou od auta táto miera klesá.

Činnosť kráčania dvoch a viacerých ľudí je spojením vzťahu medzi časom, v ktorom sa ľudia vyskytovali v rámci videa a porovnania vzdialenosťami medzi ich trajektóriami. Je možné určiť, že pokiaľ sa človek A vyskytoval v časovom intervale $\langle t, t + n \rangle$ a človek B v intervale $\langle t + m, t + o \rangle$, pre ktoré platí že $m < n$ a ich vzdialenosť Fréchet je menšia ako určitá stanovená hranica, tak sa tieto dve osoby v stanovenom intervale $\langle t + m, t + n \rangle$

pohybovali spolu. Zároveň pokiaľ sa ďalšia osoba C pohybovala spolu s osobou B v intervale $\langle t+p, t+q \rangle$, pre ktorý platí že $p < m$, $p < n$ a $q > n$, je možné určiť že aj osoba C sa pohybovala s osobou A a B , avšak iba v intervale $\langle t+p, t+n \rangle$ (obrázok 6.9). Miera istoty sa v prípade tejto činnosti určuje z aritmetického priemeru hodnôt miery istoty rozpoznania všetkých osôb zahrnutých v danej činnosti kráčania spolu po dobu trvania akcie a aritmetického priemeru vzdialenosí Fréchet medzi nimi.



Obr. 6.9: Znázornenie intervalov výskytu jednotlivých objektov

Pre vyhľadanie činnosti presúvania sa medzi viacerými definovanými miestami, je potrebné najprv tieto miesta zadefinovať. Miesta sú predstavené polygónom, ktorý musí byť v tvare po sebe idúcich bodov a musí obsahovať uzatvárajúci bod. Definícia jednotlivých polygónov je v podobe *JSON*⁴ zápisu, ktorý je možné vidieť na obrázku 6.10. Následne sú po spustení vyhľadávania získané výsledky v podobe objektov, ktoré sa presúvali medzi miestami v poradí akom boli zadefinované. Pre objekty platí pravidlo, že sa nachádzajú v definovanej oblasti pokiaľ sa v nej nachádzajú ich stredové body.

```
{
  "polygons": [
    [[7, 765], [328, 575], [432, 553], [42, 844], [7, 765]],
    [[1053, 537], [605, 537], [555, 585], [1045, 590], [1053, 537]]
  ]
}
```

Obr. 6.10: Ukážka *JSON* notácie zápisu definície miest, medzi ktorými má byť vyhľadávaný pohyb

Miera istoty je v tomto prípade odvodená jednoduchým spôsobom, a to výpočtom aritmetického priemeru miery istoty rozpoznania daného objektu, ktorý prechádza jednotlivými definovanými oblasťami videa.

⁴JavaScript Object Notation

Kapitola 7

Testovanie

Testovanie implementovaného systému bolo vykonané pomocou troch prístupov. Prvé testy sú zamerané na rýchlosť spracovania a analýzy videa. Ďalšie testujú spracovávanie jednotlivých informácií a ukladanie do databázy. Pri tomto testovaní bola taktiež otestovaná rýchlosť. Nakoniec v poslednom teste figurujú definované činnosti ľudí pomocou vzťahov medzi získanými informáciami. V tomto teste je sledovaná presnosť a schopnosť rozpoznávania definovaných činností. Rýchlosť testov sú plne závislé na hardvéri, na ktorom boli vykonávané.

Pre testovanie definovaných činností bola vybratá dátová sada *VIRAT Video Dataset* [39], ktorá sa zameriava práve na udalosti spomenuté v sekcií 6.6 a niektoré ďalšie iné. Pred použitím tejto dátovej sady bolo potrebné získať úseky videa, na ktorých sa jednotlivé činnosti ľudí nachádzajú, keďže táto dátová sada obsahuje videá rôznych dĺžok s viacerými udalosťami vo videách. Činnosti boli z videí vystrihnuté tak, aby bola v testovacom videu celá vykonávaná činnosť, spolu s krátkym časovým úsekom po jej vykonaní.

Analýza a spracovávanie informácií bolo vykonávané pomocou vysokovýkonnej grafickej karty *Nvidia RTX 2080Ti*, ktorá zrýchľuje tento proces. Táto rýchlosť je taktiež ovplyvňovaná rôznymi ďalšími faktormi ako sú:

- rozlíšenie videa,
- vzorkovacia rýchlosť videa,
- počet vyskytujúcich sa objektov.

Práve z toho dôvodu boli testovacie videá rozdelené do dvoch skupín, a to podľa rozlíšenia a vzorkovacej rýchlosťi videa uvádzanej v obrázkoch za sekundu (z angl. frames per second – fps). Kedže nie je vopred známe, koľko objektov sa v testovacích videách nachádza, skupina podľa počtu objektov nebola vytvorená.

Implementácia systému využíva funkciu preskakovania obrázkov, vďaka ktorej je rozpoznávanie informácií z videa rýchlejšie. Preskakovanie obrázkov zabezpečuje, že na danej sekunde videa prebehne proces analýzy iba niekoľkokrát, čím je zvýšená rýchlosť a zachovanie informatívnej hodnoty. Avšak videá s rôznymi rýchlosťami vzorkovania obrázkov za sekundu na túto implementáciu odpovedajú inými rýchlosťami spracúvania. Preto bude výsledná rýchlosť analýzy vyjadrená pomocou pomeru medzi vzorkovacou rýchlosťou videa a rýchlosťou spracúvania.

Pri testovaní rýchlosťi analyzovania informácií z videa bolo dokázané, že pre spracovanie videa s rozlíšením menším alebo rovným 1920×1080 , je systém schopný rozpoznať infor-

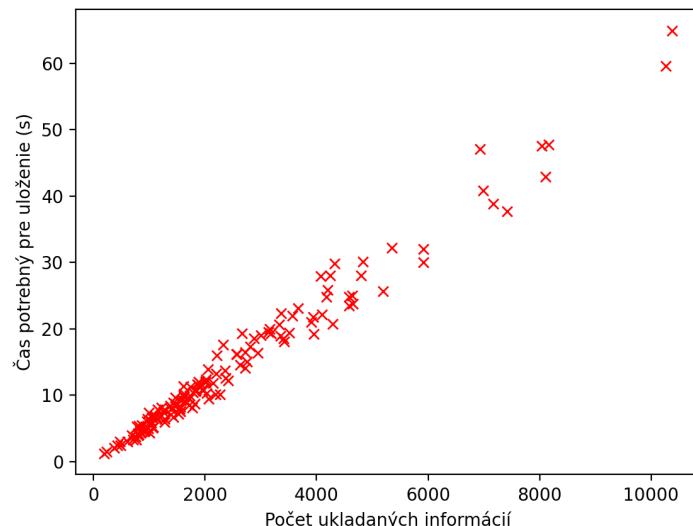
Vzorkovacia rýchlosť	Rýchlosť spracúvania	Rozlíšenie	$\frac{\text{Rýchlosť spracúvania}}{\text{Vzorkovacia rýchlosť}}$	Počet videí
30 fps	29.9 fps	1920×1080	0.9993	78
30 fps	43.5 fps	1280×720	1.4479	39
24 fps	33.3 fps	1280×720	1.3882	25

Tabuľka 7.1: Výsledky testovanie rýchlosť analýzy podľa rozdelenia rozlíšenia a vzorkovacej rýchlosť videa

mácie v reálnom čase (obrázok 7.1). Zároveň testovanie ukázalo, že čím menším rozlíšením video disponuje, tým sa čas potrebný pre analyzovanie skracuje.

Čas ukladania dát je z hľadiska testovania na krátkych videách zanedbateľným faktorom. Avšak pokiaľ je systém nasadený na videozáZNAMY obsahujúce obrovské množstvo informácií, ukladanie dát by dokázalo zabrať desiatky minút až niekoľko hodín. Testovanie rýchlosť ukladania dát bolo vykonané na školskom servery *Athena1*, na ktorom je nainštalovaná *Postgres* databáza spolu s geografickým rozšírením *Postgis*.

Vzhľadom na to, že pre testovanie bolo použitých mnoho videí, nie je možné o každom uviesť ako dlho trvalo ukladanie informácií. Taktiež by táto rýchlosť bola ovplyvňovaná počtom rozpoznaných informácií v konkrétnom videu. Z obrázku 7.1 je možné vidieť, ako sa pri testovaní jednotlivé časy ukladania menili v závislosti na počte ukladaných informácií.



Obr. 7.1: Zobrazenie výsledných časov v závislosti na počte ukladaných dát

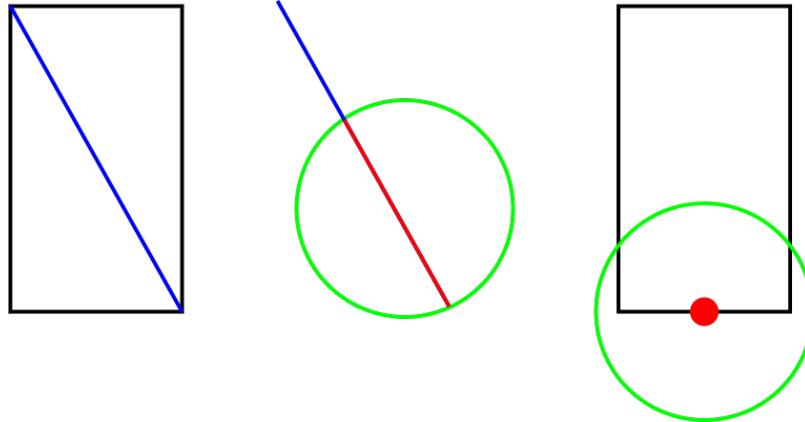
Výsledky testovania ukázali, že ukladanie vykazuje lineárnu závislosť medzi počtom ukladaných dát a časom potrebným pre ich uloženie. Priemerná rýchlosť uloženia tisíc jednotiek informácií je 5.733 sekúnd. Táto hodnota bola nameraná z priemernej hodnoty každého testovaného videa.

Posledným a najpodstatnejším testovaním systému je schopnosť rozpoznávať činnosti pomocou definovania vzťahov medzi informáciami. Pre testovanie úspešnosti rozpoznávania boli definované činnosti, spomenuté v sekcii 6.6.

Pred každým testom na nejakom videu, musela byť najprv vykonaná analýza videa. Prvým krokom pri tomto testovaní bolo získanie súborov s uloženými informáciami o jednotlivých objektoch v rámci videa. Po ich získaní boli súbory postupne v cykle spracované a ukladané do databázy. Následne bol spúštaný skript s definíciou vzťahov medzi uloženými informáciami, ktorý vyhľadával úsek videa s požadovanou činnosťou.

Každá z činností disponuje určitými parametrami, ktoré musia byť nastavené pre správne rozpoznanie činnosti. Pri činnostiach nastupovania a vystupovania z auta je zásadné nastaviť, vzdialenosť medzi prvým a posledným bodom trajektórie a maximálnu vzdialenosť od auta. Taktiež je pre obidve navrhnutá hodnota minimálneho trvania výskytu človeka. Pri činnosti kráčania dvoch a viacerých ľudí spolu sú týmito nastaviteľnými parametrami maximálny uhlový rozdiel určujúci rozdiel smerov kráčania dvoch osôb, maximálny rozdiel rýchlosťí dvoch osôb a maximálnu hodnotu vzdialenosťi *Fréchet* medzi dvoma objektmi. Posledná definovaná činnosť, presúvanie sa medzi viacerými miestami, nedisponuje žiadnymi nastaviteľnými parametrami.

Pri nastavovaní parametrov činnosti nastupovania do auta je potrebné si uvedomiť, ako jednotlivé informácie spolu súvisia. Aby bola akcia úspešne rozpoznaná, nesmie byť posledný bod trajektórie človeka príliš vzdialený od auta. Túto hodnotu je však pre každé video nutné nejakým spôsobom vypočítať. Kedže je vzdialenosť pre každé video rozdielna z dôvodu rôznych scén a veľkostí objektov, je táto hodnota nastavovaná pomocou diagonály ohraničujúceho boxu človeka, u ktorého sa rozhoduje o vykonávanej činnosti (obrázok 7.2). Ďalšia nastaviteľná hodnota vzdialenosťi prvého a posledného bodu v trajektórii je hodnota, ktorá je nastavovaná pre korekciu detekcií. V niektorých prípadoch model pre sledovanie objektov nie je schopný priradiť človeku jednotný identifikátor, a preto táto hodnota slúži pre korekciu falošne pozitívnych výsledkov.



Obr. 7.2: Grafický príklad výpočtu maximálnej vzdialenosťi (červená čiara) od auta podľa diagonály ohraničujúceho boxu (modrá čiara) človeka (pokiaľ zelený kruh pretne ohraničujúci box auta, ide o pozitívnu detekciu)

Pri činnosti vystupovania z auta sa nastavujú rovnaké parametre ako pri nastupovaní. Obidve hodnoty maximálnej vzdialenosťi od auta a minimálna vzdialenosť medzi prvým a posledným bodom trajektórie sú vypočítané pomocou diagonály ohraničujúceho boxu človeka. Poslednou spomínanou nastaviteľnou hodnotou je minimálna dĺžka výskytu človeka vo videu. Táto hodnota určuje minimálny počet detekcií rozpoznaných systémom pre objekt človeka, aby išlo o pozitívne rozpoznanie činnosti.

Činnosť kráčania dvoch alebo viacerých ľudí spolu je závislá od nastavenia dvoch parametrov. Maximálnej vzdialenosť *Fréchet* medzi dvomi ľuďmi a maximálneho rozdielu aritmetického priemeru ich smerov kráčania. Maximálnu hodnotu vzdialenosť *Fréchet* je rovako ako pri predchádzajúcich činnostach nutné nastaviť pre každé video rozdielne. Pre jej nastavenie je použitý aritmetický priemer dĺžok diagonál všetkých ohraničujúcich boxov ľudí v každom čase výskytu vo videu. Táto hodnota je upravovaná na základe rôznych násobkov. Druhou nastaviteľnou hodnotou je maximálna veľkosť rozdielu medzi smermi ľudí, ktorá je nastavovaná tak, aby celkový rozdiel smerov dvoch ľudí nepresiahol 90 stupňov.

Všetky tieto nastaviteľné hodnoty ovplyvňujú úspešnosť rozpoznania činností. V nasledujúcich tabuľkách 7.2, 7.3 a 7.4 sú zobrazené najlepšie dosiahnuté výsledky testov, spolu s výslednými hodnotami pravdivo pozitívnych (PP), falošne pozitívnych (FP) a falošne negatívnych (FN) rozpoznaní, po nastavení jednotlivých parametrov. Testy sú vyhodnotené pomocou metriky priemernej presnosti vysvetlenej v sekcií 3.4. Ako bolo spomenuté vyššie, minimálne a maximálne vzdialenosť sú vyjadrené pomocou násobkov dĺžok diagonál ohraničujúcich boxov ľudí. Vyhodnotenie poslednej činnosti presúvanie sa medzi definovanými miestami je zobrazené v tabuľke 7.5.

Minimálna vzdialenosť medzi prvým a posledným bodom trajektórie	Maximálna vzdialenosť od auta	Minimálny počet detekcií	Priemerná presnosť	PP/FP/FN
0.3	0.6	12	90.2326%	42/18/7
0.2	0.4	12	87.5853%	42/18/7
0.2	0.2	12	85.5290%	38/13/11
0.1	0.2	10	83.8290%	41/15/8

Tabuľka 7.2: Výsledky testovania parametrov definovanej činnosti nastupovanie do auta (testované na 49 prípadoch)

Definovaná činnosť nastupovania do auta, dokázala rozpoznať väčšinu prípadov vyskytujúcich sa vo videách. Výsledok testovania zároveň však ukazuje vysoký počet falošne pozitívnych detekcií. Značný počet týchto detekcií pochádza z toho, že definícia nastupovania do auta počíta s koncom trajektórie človeka pri aute. Preto je činnosť chybne rozpoznávaná aj v prípadoch keď iný človek prechádza okolo auta v momente kedy videozáZNAM skončí.

Minimálna vzdialenosť medzi prvým a posledným bodom trajektórie	Maximálna vzdialenosť od auta	Minimálny počet detekcií	Priemerná presnosť	PP/FP/FN
0.6	0.6	12	89.5618%	27/13/25
0.2	0.4	10	89.2249%	29/11/23
0.3	0.6	12	87.4144%	28/18/24
0.2	0.6	12	86.5963%	28/18/24

Tabuľka 7.3: Výsledky testovania parametrov definovanej činnosti vystupovanie z auta (testované na 52 prípadoch)

Rozpoznávanie vystupovania z auta trápi rovnaký problém ako pri nastupovaní do auta, no v tomto prípade ide o hraničné prípady na začiatku videozáZNAMOV, pokiaľ sa človek

nachádza v blízkosti auta. Tieto problémy pri činnostiach by boli nasadením na dlhých videozáznamoch z veľkej časti eliminované.

Maximálny rozdiel uhlov smeru dvoch ľudí	Maximálna vzdialenosť dvoch trajektórií	Priemerná presnosť	PP/FP/FN
70	1.5	89.3229%	43/6/13
70	2.0	84.3899%	47/12/9
90	2.0	78.5877%	52/22/4
90	2.5	64.5525%	49/37/7

Tabuľka 7.4: Výsledky testovania parametrov definovanej činnosti kráčania dvoch a viacerých ľudí spolu (testované na 56 prípadoch)

Definícia kráčania dvoch a viacerých ľudí rozpoznáva činnosti s najvyššou mierou úplnosti. Dokáže rozpoznať takmer všetky prípady, kedy ľudia išli spolu, avšak z dôvodu nastavenia maximálnej vzdialenosťi dvoch trajektórií trpí množstvom falošne pozitívnych detekcií. Tento problém sa vyskytuje najmä pri videozáznamoch, v ktorých sa objavujú osoby v príliš rozdielnych vzdialostiah od bezpečnostnej kamery.

Pravdivo pozitívne	Falošne pozitívne	Falošne negatívne	F1 skóre
29	0	10	85.2941%

Tabuľka 7.5: Výsledky testovania definovanej činnosti presúvania sa medzi viacerými mestami (testované na 39 prípadoch)

Definovaná činnosti prechádzania medzi viacerými mestami vykázala úplné obmedzenie falošne negatívnych detekcií. Avšak stále pretrváva problém s rozpoznaním všetkých vyskytujúcich sa situácií. V takomto prípade by priemerná presnosť dosahovala najvyššiu možnú hodnotu, čo nevypovedá vhodne o schopnosti rozpoznávať činnosť. Preto bola pôvodná vyhodnocovacia metrika zamenená za metriku F1 skóre.

Pri testovaní systémov pre rozpoznavanie činností ľudí by bolo možné zohľadniť aj náročnosť situácií vyskytujúcich sa vo videu. Vzhľadom na veľké rozdiely v náročnostiach rozpoznania činností medzi kamerovým záznamom s vysokým rozlíšením, na ktorom sa vyskytujú dostatočne viditeľné a veľké objekty, ktoré je možné rozpoznať pomocou navrhnutých modelov pomerne ľahko a kamerovým záznamom s nižším rozlíšením a vzdialenosťmi objektami, s ktorými majú modely väčšie problémy, by sa dali jednotlivé videá, na ktorých je systém testovaný, zaradiť do skupín podľa náročnosti na základe určitých kritérií. Ďalším rozdelením videí, ktoré by pri testovaní mali byť zohľadnené, sú prípady týkajúce sa konkrétnych činností. Napríklad pri rozpoznavaní činností nastupovania a vystupovania z auta je faktorom ovplyvňujúcim náročnosť rozpoznania napríklad strana, z ktorej človek činnosť vykonáva. Pokiaľ sa človek nachádza za autom, model nie je schopný rozpoznať takúto situáciu, pretože ho nevidí.

Pri vývoji systémov pre prehľadávanie videí sú testovacie videá často hrané, alebo sú sprostredkované od objednávateľov systémov. Toto je často zapríčinené etickými a právnymi problémami spojenými so získavaním videozáznamov. Zároveň sú systémy vyvíjané k účelu priamo spätemu s požiadavkami objednávateľov, a preto zabezpečujú videá oni. Z toho dôvodu nebolo možné zostaviť dostatočne veľkú testovaciu sadu, ktorá by sa dala rozdeliť na rôzne skupiny podľa náročností jednotlivých činností. Pri rozdelení už takto obmedzenej dátovej sady by testovanie dokázal zmeniť už jeden nepravdivo pozitívny výsledok na toľko,

že by mohli z analýzy výsledkov vyjsť nepravdivé závery. Testovanie bolo preto vykonané na dátovej sade ako celku. V tomto prípade však nevieme povedať, ako by sa systém zachoval pri takýchto rôzne náročných skupinách videí.

Testovanie ukázalo, že takáto definícia vzťahov medzi jednotlivými informáciami má veľmi dobrú priemernú presnosť aj vzhľadom na to, že neboli rozpoznané všetky akcie vyskytujúce sa vo videách. Avšak problémy s rozpoznávaním jednotlivých činností podliehajú viacerým faktorom. Jedným z dôvodov neschopnosti rozpoznať jednotlivé akcie môže byť samotné získavanie informácií, ktoré je ovplyvňované výzvami rozpoznávania a sledovania objektov. V tomto prípade by mohli byť zvolené iné modely pre časti systému, čo by však ovplyvnilo schopnosť rozpoznať a spracovať informácie v reálnom čase.

Kapitola 8

Záver

Hlavným cieľom tejto práce bolo navrhnuť a vytvoriť systém schopný rozpoznávať ľudí a ich činností z bezpečnostných kamier. Cieľ práce bol splnený pomocou navrhnutia a implementácie systému schopného rozpoznávať objekty, sledovať ich a zároveň pre objekty typu človek rozpoznať jednoduché typy akcií. Na základe získaných informácií o objektoch boli definované vzťahy medzi nimi, ktoré predstavujú činnosti ľudí rozpoznateľné systémom. Následným vyhľadávaním týchto vzťahov v databáze s uloženými informáciami bol systém schopný rozpoznávať požadované úseky videozáznamov, ktorým bola priradená miera istoty určujúca hodnotu presvedčenia, že sa v danom úseku definovaná činnosť nachádza.

Boli vykonané počiatocné experimenty na existujúcich modeloch schopných rozpoznávať objekty, na základe ktorých bol spomedzi všetkých modelov vybraný *YOLO* štvrtej verzie, ktorý dosahoval najvyššiu rýchlosť pri zachovaní vhodnej presnosti. Pri týchto experimentoch bolo dokázané že pri rozlíšení videa 1920×1080 je tento model schopný rozpoznávať objekty rýchlosťou 19 obrázkov za sekundu, s vysokou presnosťou, kedy mal najmenší rozpoznaný objekt rozmery 19×44 pixelov, čím prekonal ostatné modely. Táto rýchlosť bola naviac zvýšená pomocou výberu relevantného počtu obrázkov za sekundu, nad ktorými model vykonával rozpoznávanie.

Pri sledovaní objektov bol najväčší dôraz kladený na schopnosť sledovania objektov v reálnom čase, aby nebol obmedzovaný a spomalený model pre rozpoznávanie objektov. Preto bol pre tento účel vybraný algoritmus *Deep SORT*, ktorý túto rýchlosť dosahoval spolu s presnosťou porovnatelnou s ostatnými algoritmami. Do algoritmu bol doplnený údaj o smere a rýchlosťi sledovaných objektov v rámci obrázku.

Rozpoznávanie informácií o vykonávaní jednoduchých akcií ľudí, akými sú napríklad kráčanie alebo nosenie predmetov, bolo zabezpečené pomocou modelu *SlowFast*, ktorý predčil všetky dosiaľ vytvorené modely neurónových sietí v presnosti aj rýchlosťi [16].

Medzi informáciami, ktoré je systém schopný rozpoznávať, boli definované vzťahy predstavujúce činnosti ľudí, akými je nastupovanie do auta, vystupovanie z auta, kráčanie dvoch a viacerých ľudí spolu a presúvanie sa medzi definovanými miestami. Následne bolo na dátovej sade *VIRAT Video Dataset* vykonané testovanie výsledného systému pre schopnosť rozpoznávať definované činnosti. Prvým krokom bola úprava a prehľadávanie dátovej sady pre výber relevantných úsekov videí. Pre účel testovania bolo vystrihnutých niekoľko desiatok videí s činnosťami nastupovanie do auta, vystupovanie z auta, kráčanie dvoch a viacerých ľudí spolu a presúvanie sa medzi definovanými miestami.

Už prvé výsledky testovania ukázali vysoké percento priemernej presnosti, kedy nebola u žiadnej z činností nameraná hodnota menšia ako 60%. Pre definované vzťahy medzi získanými informáciami predstavujúce činnosti boli testované rôzne nastavenia parametrov a

najlepšie nastavenia ukázali, že rozpoznanie činnosti nastupovania do auta vykazuje hodnotu 90.23% priemernej presnosti. Ďalšie činnosti, vystupovanie z auta a kráčanie dvoch a viacerých ľudí spolu dosiahli rovnako dobré výsledky, kedy prvá spomenutá dosiahla 89.56% a druhá 89.32% priemernej presnosti. Posledná činnosť presúvanie sa medzi definovanými miestami bola testovaná metrikou F1 skóre, vzhladom na absenciu falošne pozitívnych detekcií. Táto metrika vykázala hodnotu skóre 85.29%.

Z hľadiska budúceho vývoja by bolo zaujímavým rozšírením systému návrh a implementácia užívateľského rozhrania, ktoré by uľahčilo používanie koncovému užívateľovi. Pomocou takéhoto rozhrania by bolo možné vkladať jednotlivé videozáznamy do systému, bez nutnosti poznat náročné príkazy a techniky spúšťania pomocou príkazového riadka.

Ďalším vhodným rozšírením systému by bolo vytvorenie popisujúceho jazyka, definujúceho vzťahy medzi informáciami. Jeho pomocou by bolo možné skladať konštrukcie predstavujúce jednoduché činnosti, ktoré by sa dali rozširovať a spájať s ďalšími konštrukciami, na základe čoho by bolo uľahčené definovanie zložitejších a komplexnejších činností.

Literatúra

- [1] ALBAWI, S., MOHAMMED, T. A. a AL ZAWI, S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)* [online]. IEEE, 2017, s. 1–6 [cit. 2021-03-31]. DOI: 10.1109/ICEngTechnol.2017.8308186. ISBN 978-1-5386-1949-0. Dostupné z: <https://ieeexplore.ieee.org/document/8308186/>.
- [2] ANTON, M., LEAL TAIXE, L., REID, I. et al. MOT16: A Benchmark for Multi-Object Tracking. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2016, abs/1603.00831, 3.5.2016, [cit. 2021-02-28]. Dostupné z: <https://arxiv.org/abs/1603.00831>.
- [3] ARONOV, B., HAR PELED, S., KNAUER, C. et al. Fréchet Distance for Curves, Revisited. *CoRR* [online]. April 2015, abs/1504.07685, [cit. 2021-05-06]. Dostupné z: <http://arxiv.org/abs/1504.07685>.
- [4] ARROYO, R., YEBES, J. J., BERGASA, L. M. et al. Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert systems with applications* [online]. Elsevier Ltd. 2015, zv. 42, č. 21, s. 7991–8005, [cit. 2021-04-30]. DOI: 10.1016/j.eswa.2015.06.016. ISSN 0957-4174. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0957417415004182>.
- [5] BEBIS, G., EGBERT, D. a SHAH, M. Review of computer vision education. *IEEE Transactions on Education* [online]. 2003, zv. 46, č. 1, s. 2–21, [cit. 2021-03-19]. DOI: 10.1109/TE.2002.808280. ISSN 0018-9359. Dostupné z: <http://ieeexplore.ieee.org/document/1183662/>.
- [6] BERGMANN, P., MEINHARDT, T. a LEAL TAIXE, L. Tracking without bells and whistles. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2019, abs/1903.05625, 17.8.2019, [cit. 2021-02-06]. Dostupné z: <https://arxiv.org/abs/1903.05625>.
- [7] BEWLEY, A., GE, Z., OTT, L. et al. Simple online and realtime tracking. In: *2016 IEEE International Conference on Image Processing (ICIP)* [online]. IEEE, 2016, s. 3464–3468 [cit. 2021-03-27]. DOI: 10.1109/ICIP.2016.7533003. ISBN 978-1-4673-9961-6. Dostupné z: <http://ieeexplore.ieee.org/document/7533003/>.
- [8] BOCHKOVSKIY, A., CHIEN YAO, W. a HONG YUAN, M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2020, abs/2004.10934, [cit. 2021-03-25]. Dostupné z: <https://arxiv.org/abs/2004.10934>.

- [9] CHINCHOR, N. MUC-4 Evaluation Metrics. In: USA: Association for Computational Linguistics, 1992, s. 22–29. MUC4 '92. DOI: 10.3115/1072064.1072067. ISBN 1558602739. Dostupné z: <https://doi.org/10.3115/1072064.1072067>.
- [10] COROVIC, A., ILIC, V., DURIC, S. et al. The Real-Time Detection of Traffic Participants Using YOLO Algorithm. In: *2018 26th Telecommunications Forum (TELFOR)* [online]. IEEE, 2018, s. 1–4 [cit. 2021-05-02]. DOI: 10.1109/TELFOR.2018.8611986. Dostupné z: <https://ieeexplore.ieee.org/document/8611986>.
- [11] DANIELSSON, P.-E. Euclidean distance mapping. *Computer Graphics and Image Processing* [online]. Október 1980, zv. 14, č. 3, s. 227–248, [cit. 2021-05-06]. DOI: 10.1016/0146-664X(80)90054-4. ISSN 0146-664X. Dostupné z: <https://www.sciencedirect.com/science/article/pii/0146664X80900544>.
- [12] DEORI, B. a THOUNAOJAM, D. A Survey on Moving Object Tracking in Video. *International Journal on Information Theory* [online]. Júl 2014, zv. 3, s. 31–46, [cit. 2021-05-06]. DOI: 10.5121/ijit.2014.3304. Dostupné z: https://www.researchgate.net/publication/283896922_A_Survey_on_Moving_Object_Tracking_in_Video.
- [13] DEVASENA, C., REVATHÍ, R. a HEMALATHA, M. Video Surveillance Systems - A Survey. *International Journal of Computer Science Issues (IJCSI)* [online]. Mahebourg: International Journal of Computer Science Issues (IJCSI). 2011, zv. 8, č. 4, s. 635–642, [cit. 2021-04-30]. ISSN 1694-0814. Dostupné z: <http://www.ijcsi.org/papers/IJCSI-8-4-1-635-642.pdf>.
- [14] DU, Y., WANG, W. a WANG, L. Hierarchical recurrent neural network for skeleton based action recognition. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [online]. IEEE, 2015, s. 1110–1118 [cit. 2021-03-31]. DOI: 10.1109/CVPR.2015.7298714. ISBN 978-1-4673-6964-0. Dostupné z: <https://ieeexplore.ieee.org/document/7298714/>.
- [15] EITER, T. a MANNILA, H. Computing discrete Fréchet distance. Citeseer, apríl 1994 [cit. 2021-05-05]. Dostupné z: https://www.researchgate.net/publication/228723178_Computing_Discrete_Frechet_Distance.
- [16] FEICHTENHOFER, C., FAN, H., MALIK, J. a HE, K. SlowFast Networks for Video Recognition. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2018, abs/1812.03982, 29.10.2019, [cit. 2021-02-27]. Dostupné z: <https://arxiv.org/abs/1812.03982>.
- [17] GU, C., CHEN, S., ROSS, D. et al. AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2017, abs/1705.08421, [cit. 2021-02-27]. Dostupné z: <https://arxiv.org/abs/1705.08421>.
- [18] GUNAWARDANA, A. a SHANI, G. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *J. Mach. Learn. Res.* [online]. JMLR.org. December 2009, zv. 10, s. 2935–2962, [cit. 2021-05-06]. ISSN 1532-4435. Dostupné z: <https://dl.acm.org/doi/10.5555/1577069.1755883>.

- [19] HECKBERT, P. S. *Graphics gems IV*. Boston, USA: Academic Press Professional, Inc., august 1994 [cit. 2021-04-05]. 24-46 s. ISBN 0-12-336155-9.
- [20] HELD, C., KRUMM, J., MARKEL, P. et al. Intelligent Video Surveillance. *Computer* [online]. IEEE. 2012, zv. 45, č. 3, s. 83–84, [cit. 2021-04-30]. DOI: 10.1109/MC.2012.97. ISSN 0018-9162. Dostupné z: <https://ieeexplore.ieee.org/document/6163452>.
- [21] HENDERSON, P. a FERRARI, V. End-to-end training of object class detectors for mean average precision. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2016, abs/1607.03476, 16.3.2017, [cit. 2021-02-05]. Dostupné z: <https://arxiv.org/abs/1607.03476>.
- [22] HERATH, S., HARANDI, M. a PORIKLI, F. Going deeper into action recognition: A survey. *Image and vision computing* [online]. Elsevier B.V. 2017, zv. 60, C, s. 4–21, [cit. 2021-04-02]. DOI: 10.1016/j.imavis.2017.01.010. ISSN 0262-8856. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0262885617300343>.
- [23] HUANG, T. Computer vision: Evolution and promise. [online]. Cern. 1996, s. 21–25, [cit. 2021-03-01]. DOI: 10.5170/CERN-1996-008.21. Dostupné z: <http://cds.cern.ch/record/400313>.
- [24] JEGHAM, I., KHALIFA, A. B., ALOUANI, I. et al. Vision-based human action recognition: An overview and real world challenges. *Forensic Science International: Digital Investigation* [online]. Elsevier Ltd. 2020, zv. 32, [cit. 2021-04-05]. DOI: 10.1016/j.fsid.2019.200901. ISSN 2666-2817. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S174228761930283X>.
- [25] KAEHLER, A. a BRADSKI, G. R. *Learning OpenCV 3: Computer Vision in C++ with the OpenCV library*. Sebastopol: O'Reilly, 2016 [cit. 2021-03-20]. ISBN 978-1-4919-3799-0.
- [26] KAY, W., CARREIRA, J., SIMONYAN, K. et al. The Kinetics Human Action Video Dataset. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2017, abs/1705.06950, 30.4.2017, [cit. 2021-02-28]. Dostupné z: <https://arxiv.org/abs/1705.06950>.
- [27] KONG, Y. a FU, Y. Human Action Recognition and Prediction: A Survey. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2018, abs/1806.11230, 2.7.2018, [cit. 2021-04-05]. Dostupné z: <https://arxiv.org/abs/1806.11230>.
- [28] KUMAR, R., HIRVONEN, D., HANSEN, M. et al. Aerial video surveillance and exploitation. *Proceedings of the IEEE* [online]. IEEE. 2001, zv. 89, č. 10, s. 1518–1539, [cit. 2021-05-01]. DOI: 10.1109/5.959344. ISSN 0018-9219. Dostupné z: <https://ieeexplore.ieee.org/document/959344>.
- [29] LECUN, Y., KAVUKCUOGLU, K. a FARABET, C. Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems* [online]. IEEE, 2010, s. 253–256 [cit. 2021-03-05]. DOI: 10.1109/ISCAS.2010.5537907. ISBN 978-1-4244-5308-5. Dostupné z: <http://ieeexplore.ieee.org/document/5537907/>.

- [30] LEI, Q., DU, J.-x., ZHANG, H.-b. et al. A Survey of Vision-Based Human Action Evaluation Methods. *Sensors (Basel, Switzerland)* [online]. 2019, zv. 19, č. 19, [cit. 2021-05-02]. DOI: 10.3390/s19194129. Dostupné z: <https://www.mdpi.com/1424-8220/19/19/4129>.
- [31] LI, C., HOU, Y., WANG, P. et al. Joint Distance Maps Based Action Recognition With Convolutional Neural Networks. *IEEE Signal Processing Letters* [online]. 2017, zv. 24, č. 5, s. 624–628, [cit. 2021-03-27]. DOI: 10.1109/LSP.2017.2678539. ISSN 1070-9908. Dostupné z: <http://ieeexplore.ieee.org/document/7872453/>.
- [32] LI, X., HU, W., SHEN, C. et al. A survey of appearance models in visual object tracking. *ACM Transactions on Intelligent Systems and Technology* [online]. 2013, zv. 4, č. 4, s. 1–48, [cit. 2021-03-31]. DOI: 10.1145/2508037.2508039. ISSN 2157-6904. Dostupné z: <https://dl.acm.org/doi/10.1145/2508037.2508039>.
- [33] LIU, L., WANG, X., FIEGUTH, P. et al. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision* [online]. New York: Springer Nature B.V. 2020, zv. 128, č. 2, s. 261–318, [cit. 2021-05-02]. DOI: 10.1007/s11263-019-01247-4. ISSN 09205691. Dostupné z: <https://link.springer.com/article/10.1007/s11263-019-01247-4>.
- [34] LIU, T., FANG, S., ZHAO, Y. et al. Implementation of Training Convolutional Neural Networks. *CoRR* [online]. Jún 2015, abs/1506.01195, 4.6.2015, [cit. 2021-03-04]. Dostupné z: <http://arxiv.org/abs/1506.01195>.
- [35] LU, X., LI, Q., LI, B. et al. MimicDet: Bridging the Gap Between One-Stage and Two-Stage Object Detection. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2020, abs/2009.11528, [cit. 2021-04-03]. Dostupné z: <https://arxiv.org/abs/2009.11528>.
- [36] LUO, W., XING, J., ANTON, M. et al. Multiple Object Tracking: A Literature Review. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2014, abs/1409.7618, 22.4.2017, [cit. 2021-02-04]. Dostupné z: <https://arxiv.org/abs/1409.7618>.
- [37] MABROUK, A. a ZAGROUBA, E. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications* [online]. New York: Elsevier BV. 2018, zv. 91, s. 480, [cit. 2021-04-30]. ISSN 0957-4174. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0957417417306334>.
- [38] MAHTO, P., GARG, P., SETH, P. et al. Refining Yolov4 for Vehicle Detection. *International Journal of Advanced Research in Engineering and Technology (IJARET)* [online]. Jún 2020, zv. 11, č. 5, s. 409–419, [cit. 2021-03-04]. Dostupné z: <https://ssrn.com/abstract=3628439>.
- [39] OH, S., HOOGS, A., PERERA, A. et al. A large-scale benchmark dataset for event recognition in surveillance video. In: *CVPR 2011* [online]. IEEE, 2011, s. 3153–3160 [cit. 2021-04-26]. DOI: 10.1109/CVPR.2011.5995586. ISBN 9781457703942. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/5995586>.
- [40] OJHA, S. a SAKHARE, S. Image processing techniques for object tracking in video surveillance- A survey. In: *2015 International Conference on Pervasive Computing*

- (ICPC) [online]. IEEE, 2015, s. 1–6 [cit. 2021-03-31]. DOI: 10.1109/PERVASIVE.2015.7087180. ISBN 978-1-4799-6272-3. Dostupné z: <http://ieeexplore.ieee.org/document/7087180/>.
- [41] O'SHEA, K. a NASH, R. An Introduction to Convolutional Neural Networks. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2015, abs/1511.08458, 2.12.2015, [cit. 2021-04-02]. Dostupné z: <https://arxiv.org/abs/1511.08458>.
- [42] OVSENIK, L., KOLESÁROVÁ, A. a TURÁN, J. VIDEO SURVEILLANCE SYSTEMS. *Acta Electrotechnica et Informatica*. Január 2010, zv. 10, s. 46–53, [cit. 2021-05-07]. Dostupné z: https://www.researchgate.net/publication/228708805_video_surveillance_systems.
- [43] PADILLA, R., NETTO, S. a SILVA, E. da. A Survey on Performance Metrics for Object-Detection Algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)* [online]. Júl 2020, s. 237–242 [cit. 2021-05-06]. DOI: 10.1109/IWSSIP48289.2020. Dostupné z: <https://ieeexplore.ieee.org/document/9145130>.
- [44] PERROT, G., DOMAS, S. a COUTURIER, R. An optimized GPU-based 2D convolution implementation. *Concurrency and Computation: Practice and Experience* [online]. Wiley Online Library. December 2015, zv. 28, č. 16, s. 4291–4304, 30.10.2015, [cit. 2021-05-06]. DOI: 10.1002/cpe.3752. Dostupné z: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.3752>.
- [45] PINHEIRO, P. a COLLOBERT, R. Recurrent Convolutional Neural Networks for Scene Parsing. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2013, abs/1306.2795, [cit. 2021-04-04]. Dostupné z: <https://arxiv.org/abs/1306.2795>.
- [46] RAJPOOT, Q. M. a JENSEN, C. D. Video Surveillance. In: *Promoting Social Change and Democracy through Information Technology* [online]. IGI Global, 2015, s. 69–92 [cit. 2021-04-30]. DOI: 10.4018/978-1-4666-8502-4.ch004. ISBN 9781466685024. Dostupné z: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8502-4.ch004>.
- [47] REDMON, J., DIVVALA, S., GIRSHICK, R. et al. You Only Look Once: Unified, Real-Time Object Detection. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2015, abs/1506.02640, 9.5.2016, [cit. 2021-04-03]. Dostupné z: <https://arxiv.org/abs/1506.02640>.
- [48] REN, S., HE, K., GIRSHICK, R. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [online]. IEEE. 2017, zv. 39, č. 6, s. 1137–1149, [cit. 2021-04-02]. DOI: 10.1109/TPAMI.2016.2577031. ISSN 0162-8828. Dostupné z: <https://ieeexplore.ieee.org/document/7485869>.
- [49] RIEDMILLER, M. a BRAUN, H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE International Conference on Neural Networks* [online]. IEEE, 1993, s. 586–591 [cit. 2021-03-06]. DOI: 10.1109/ICNN.1993.298623. ISBN 0-7803-0999-5. Dostupné z: <http://ieeexplore.ieee.org/document/298623/>.

- [50] RUSSAKOVSKY, O., DENG, J., SU, H. et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* [online]. New York: Springer US. 2015, zv. 115, č. 3, s. 211–252, [cit. 2021-04-02]. DOI: 10.1007/s11263-015-0816-y. ISSN 0920-5691. Dostupné z: <https://link.springer.com/article/10.1007/s11263-015-0816-y>.
- [51] SASAKI, Y. The truth of the F-measure. *Teach Tutor Mater*. Október 2007. Dostupné z: https://www.researchgate.net/publication/268185911_The_truth_of_the_F-measure.
- [52] SCHATTE, P. Computing the Angle between Vectors. *Computing* [online]. Wien: Springer Verlag. 1999, zv. 63, č. 1, s. 93–96, [cit. 2021-05-06]. DOI: 10.1007/s006070050052. ISSN 0010-485X. Dostupné z: <https://link.springer.com/article/10.1007/s006070050052>.
- [53] SEVO, I. a AVRAMOVIC, A. Convolutional Neural Network Based Automatic Object Detection on Aerial Images. *IEEE Geoscience and Remote Sensing Letters* [online]. 2016, zv. 13, č. 5, s. 740–744, [cit. 2021-03-05]. DOI: 10.1109/LGRS.2016.2542358. ISSN 1545-598X. Dostupné z: <http://ieeexplore.ieee.org/document/7447728/>.
- [54] SHARMA, S. a SHARMA, S. Activation Functions in Neural Networks. *International Journal of Engineering Applied Sciences and Technology* [online]. 2020, zv. 4, č. 12, s. 310–316, [cit. 2021-04-11]. ISSN 2455-2143. Dostupné z: <https://www.ijeast.com/papers/310-316,Tesma412,IJEAST.pdf>.
- [55] SHIDIK, G. F., NOERSASONGKO, E., NUGRAHA, A. et al. A Systematic Review of Intelligence Video Surveillance: Trends, Techniques, Frameworks, and Datasets. *IEEE access* [online]. IEEE. 2019, zv. 7, s. 170457–170473, [cit. 2021-04-30]. DOI: 10.1109/ACCESS.2019.2955387. Dostupné z: <https://ieeexplore.ieee.org/document/8911368>.
- [56] SHUAI, B., BERNESHAWI, A., MODOLO, D. et al. Multi-Object Tracking with Siamese Track-RCNN. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2020, abs/2004.07786, [cit. 2021-02-06]. Dostupné z: <https://arxiv.org/abs/2004.07786>.
- [57] SONBHADRA, S. a AGARWAL, S. Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2020, abs/2005.01385, 27.4.2021, [cit. 2021-04-01]. Dostupné z: <https://arxiv.org/abs/2005.01385>.
- [58] SREENU, G. a DURAI, M. S. Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *Journal of Big Data* [online]. Cham: Springer International Publishing. 2019, zv. 6, č. 1, s. 1–27, [cit. 2021-04-30]. DOI: 10.1186/s40537-019-0212-5. Dostupné z: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0212-5>.
- [59] *Suspect sought in shooting at west Detroit gas station* [online]. [cit. 2021-05-1]. Dostupné z: <https://eu.detroitnews.com/story/news/local/detroit-city/2020/06/30/suspect-shooting-west-detroit-gas-station-project-green-light/5353462002/>.
- [60] TRAORE, B. B., KAMSU FOGUEM, B. a TANGARA, F. Deep convolution neural network for image recognition. *Ecological informatics* [online]. Elsevier B.V. 2018,

- zv. 48, s. 257–268, [cit. 2021-04-13]. DOI: 10.1016/j.ecoinf.2018.10.002. ISSN 1574-9541. Dostupné z:
<https://www.sciencedirect.com/science/article/pii/S1574954118302140>.
- [61] TSUNG YI, L., MAIRE, M., BELONGIE, S. et al. Microsoft COCO: Common Objects in Context. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2014, abs/1405.0312, 21.2.2015, [cit. 2021-04-02]. Dostupné z:
<https://arxiv.org/abs/1405.0312>.
- [62] VRAHATIS, M., MAGOULAS, G., PARSOPoulos, K. et al. Introduction to artificial neural network training and applications. [online]. Október 2000, [cit. 2021-03-10]. DOI: 10.13140/2.1.1755.2322. Dostupné z:
https://www.researchgate.net/publication/267637796_Introduction_to_artificial_neural_network_training_and_applications.
- [63] WANG, Q., ZHANG, L., BERTINETTO, L. et al. Fast Online Object Tracking and Segmentation: A Unifying Approach. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2019, abs/1703.07402, 5.5.2019, [cit. 2021-02-05]. Dostupné z:
<https://arxiv.org/abs/1703.07402>.
- [64] WILSON, D. a MARTINEZ, T. R. The general inefficiency of batch training for gradient descent learning. *Neural Networks* [online]. December 2003, zv. 16, č. 10, s. 1429–1451, [cit. 2021-05-06]. DOI: 10.1016/S0893-6080(03)00138-2. ISSN 0893-6080. Dostupné z:
<https://www.sciencedirect.com/science/article/pii/S0893608003001382>.
- [65] WOJKE, N., BEWLEY, A. a PAULUS, D. Simple online and realtime tracking with a deep association metric. In: *2017 IEEE International Conference on Image Processing (ICIP)* [online]. IEEE, 2017, 2017-, s. 3645–3649 [cit. 2021-02-23]. DOI: 10.1109/ICIP.2017.8296962. ISSN 15224880. Dostupné z:
<https://ieeexplore.ieee.org/document/8296962>.
- [66] XIAO, F., LEE, Y., GRAUMAN, K. et al. Audiovisual SlowFast Networks for Video Recognition. *ArXiv.org* [online]. Ithaca: Cornell University Library, arXiv.org. 2020, abs/2001.08740, 9.3.2020, [cit. 2021-05-05]. Dostupné z:
<https://arxiv.org/abs/2001.08740>.
- [67] XU, Z., LIU, Y., MEI, L. et al. Semantic based representing and organizing surveillance big data using video structural description technology. *The Journal of systems and software* [online]. Elsevier Inc. 2015, zv. 102, s. 217–225, [cit. 2021-04-30]. DOI: 10.1016/j.jss.2014.07.024. ISSN 0164-1212. Dostupné z:
<https://www.sciencedirect.com/science/article/pii/S0164121214001551>.
- [68] YILMAZ, A., JAVED, O. a SHAH, M. Object tracking: A survey. *ACM Computing Surveys* [online]. December 2006, zv. 38, č. 4, [cit. 2021-03-27]. DOI: 10.1145/1177352.1177355. ISSN 0360-0300. Dostupné z:
<https://dl.acm.org/doi/10.1145/1177352.1177355>.
- [69] ZAYEGH, A. a AL BASSAM, N. Neural Network Principles and Applications. In: *Digital Systems* [online]. IntechOpen, 2018-11-28 [cit. 2021-03-31]. DOI: 10.5772/intechopen.80416. ISBN 978-1-78984-540-2. Dostupné z:

<https://www.intechopen.com/books/digital-systems/neural-network-principles-and-applications>.

- [70] ZHANG, H.-b., ZHANG, Y.-x., ZHONG, B. et al. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors (Basel, Switzerland)* [online]. 2019, zv. 19, č. 5, [cit. 2021-05-02]. DOI: 10.3390/s19051005. Dostupné z: https://www.researchgate.net/publication/331394549_A_Comprehensive_Survey_of_Vision-Based_Human_Action_Recognition_Methods.
- [71] ZHIQIANG, W. a JUN, L. A review of object detection based on convolutional neural network. In: *2017 36th Chinese Control Conference (CCC)* [online]. IEEE, 2017, s. 11104–11109 [cit. 2021-03-05]. DOI: 10.23919/ChiCC.2017.8029130. ISBN 978-988-15639-3-4. Dostupné z: <http://ieeexplore.ieee.org/document/8029130/>.
- [72] ZHU, Y., LI, X., LIU, C. et al. A Comprehensive Study of Deep Video Action Recognition. *CoRR* [online]. December 2020, abs/2012.06567, [cit. 2021-05-06]. Dostupné z: <https://arxiv.org/abs/2012.06567>.

Príloha A

Obsah priloženého pamäťového média

doc Zložka obsahujúca zdrojové kódy technickej správy

text Zložka s technickou správou v elektronickej podobe

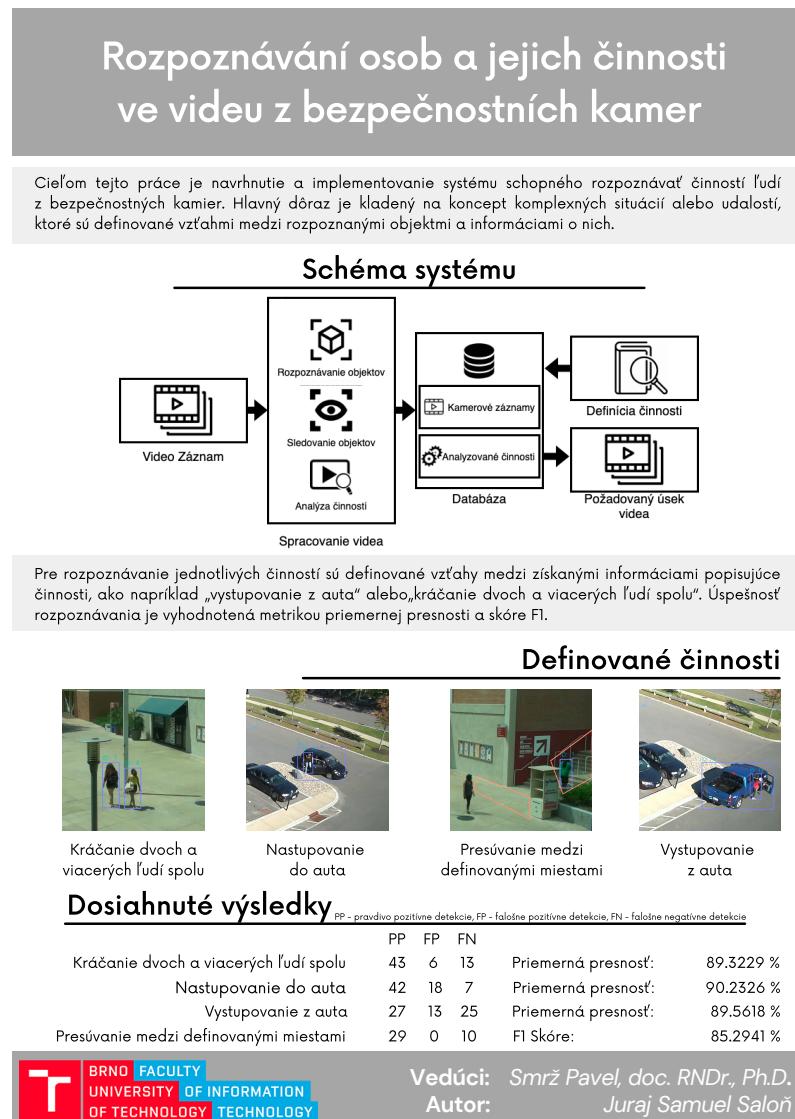
src Zložka so zdrojovými kódmi a knižnicami potrebnými k preloženiu systému

plagat.pdf Plagát prezentujúci výsledky práce a dosiahnuté ciele

README.md Textový dokument popisujúci obsah zložiek

Príloha B

Plagát



Obr. B.1: Plagát prezentujúci ciele práce a dosiahnuté výsledky.