



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

**ZPRACOVÁNÍ RADAROVÉHO SIGNÁLU
A FÚZE INFORMACÍ**

RADAR SIGNAL PROCESSING AND FUSION OF INFORMATION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

BOŘEK REICH

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Dr. Ing. PAVEL ZEMČÍK,

BRNO 2020

Zadání bakalářské práce



Student: **Reich Bořek**
Program: Informační technologie
Název: **Zpracování radarového signálu a fúze informací**
Radar Signal Processing and Fusion of Information
Kategorie: Zpracování signálů

Zadání:

1. Prostudujte dostupnou literaturu a dokumentaci k "milimetrovým radarům" a k fúzi informací z radaru a videa.
2. Navrhněte vhodnou aplikaci milimetrového radaru, s využitím dalšího senzoru, přednostně kamery, například detekci osob v místnosti, navrhněte též vhodný postup, například s využitím neuronových sítí.
3. Prostudujte možnosti implementace takové úlohy, zvažte a diskutujte vlastnosti řešení.
4. Implementujte vybranou aplikaci a demonstруйте její funkčnost na vhodném příkladu.
5. Diskutujte dosažené vlastnosti a možnosti pokračování práce.

Literatura:

- Dle pokynů vedoucího

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3 zadání

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

Vedoucí práce: **Zemčík Pavel, prof. Dr. Ing.**

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

Datum zadání: 1. listopadu 2019

Datum odevzdání: 28. května 2020

Datum schválení: 5. listopadu 2019

Abstrakt

Tato práce zkoumá možnosti fúze milimetrového radaru a kamery. Navrhuje vhodný postup a využití těchto senzorů za účelem detekce objektů, především se pak zaměřuje na vyhledávání osob a stanovení jejich vzdálenosti od senzorů. Klade důraz na použití konvolučních neuronových sítí jako prostředku pro realizaci fúze a detekce. Výstupem systému jsou informace zjištěné z dat milimetrového radaru i kamery, kdy je po provedení detekce pomocí milimetrového radaru zjištěna vzdálenost osoby. Výsledek fúze senzorů je prověřen pomocí nashromážděných dat z obou senzorů ve zhoršených světelných podmínkách, v neznámém prostředí, na neznámých osobách a pomocí referenčních údajů o vzdálenosti objektů.

Abstract

This bachelor's thesis focuses on fusion of millimeter-wave radar and camera. It proposes appropriate procedure and usage of these sensors for object detection. Object detection in this bachelor's thesis is focused on people and provides additional information about detected person. It proposes convolution neural network as means of person detection and fusion of mmWave radar and camera data. When person is detected, distance of person from sensors is found in mmWave radar point cloud. Testing is performed on input data from both sensors in different situations, in poorly lit, unknown scenes, with unknown people etc. Distance measuring is validated with reference data.

Klíčová slova

fúze informací, milimetrový radar, konvoluční neuronová síť, detekce objekt, detekce osob

Keywords

information fusion, millimeter-wave radar, convolution neural network, object detection, person detection

Citace

REICH, Bořek. *Zpracování radarového signálu a fúze informací*. Brno, 2020. Bakalářská práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce prof. Dr. Ing. Pavel Zemčík,

Zpracování radarového signálu a fúze informací

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením pana prof. Dr. Ing. Pavla Zemčíka. Uvedl jsem všechny literární prameny, publikace a další zdroje, ze kterých jsem čerpal.

.....

Bořek Reich
3. června 2020

Poděkování

Chtěl bych poděkovat svému vedoucímu bakalářské práce prof. Dr. Ing. Pavlu Zemčíkovi za odborné vedení, trpělivost a rady při zpracování této práce.

Obsah

1 Úvod	3
2 Shrnutí dosavadního stavu	4
2.1 Rozpoznávání objektů	4
2.2 Možnosti snímání objektů pomocí kamery	5
2.3 Princip fungování a způsob snímání milimetrového radaru	6
2.4 Možnosti snímání objektů za pomoci milimetrového radaru	8
2.5 Segmentace objektů za pomoci dat z kamery	10
2.6 Extrakce a výběr příznaků z dat kamery a milimetrového radaru	11
2.7 Klasifikace objektů	13
2.8 Fúze dat kamery a milimetrového radaru	20
3 Zhodnocení aktuálního stavu a návrh řešení	23
3.1 Shrnutí dosavadního stavu fúze milimetrového radaru a kamery	23
3.2 Návrh řešení fúze milimetrového radaru a kamery	23
3.3 Vhodné případy k ověření fúze milimetrového radaru a kamery pro detekci	24
3.4 Technické zadání	24
4 Popis vlastní práce	26
4.1 Sběr dat z milimetrového radaru	27
4.2 Popis hardware a umístění senzorů v prostoru	29
4.3 Transformace dat milimetrového radaru	31
4.4 Zpracování dat milimetrového radaru pro klasifikaci	33
4.5 Návrh konvoluční neuronové sítě	35
4.6 Trénování konvoluční neuronové sítě	36
4.7 Stanovení vzdálenosti detekované osoby	37
4.8 Testování systému	39
4.9 Zhodnocení výsledného řešení systému pro detekci osob	43
5 Závěr	44
Literatura	45
Seznam příloh	47
A Výsledky testovacích vstupů	48
A.1 Příklady výsledků testovacích vstupů detekce	48
A.2 Příklady výsledků test. vstupů ve zhoršených světelných podmínkách	52
B Příklady trénovacích vstupů konvoluční neuronové sítě	54

B.1	Příklady trénovacích vstupů osob	54
B.2	Příklady trénovacích vstupů prostředí	55

Kapitola 1

Úvod

V dnešní době je třeba rozeznávat osoby a jiné objekty na různých místech. Často se pro tento účel používají kamery, ty však nejsou neomylné a mohou být negativně ovlivněny prostředím, např. světelnými podmínkami, počasím apod. Dále pak není z podstaty funkčnosti kamery možné, nebo je to velmi obtížné, rozeznat od sebe určité objekty. Pro zdokonalení možností kamery je možné použít další prostředky. Tato práce se zabývá použitím radaru v kombinaci s kamerou pro zdokonalení rozpoznávání objektů.

Použití milimetrového radaru byl pro mě jeden z hlavních důvodů, proč si vybrat toto téma. Tento senzor dokáže rozeznat objekty v zákrytu, ve špatných světelných podmínkách atd. Jeho možnosti jsou navíc ne zcela prozkoumány. Kombinace radaru a kamery je ke všemu zajímavá kvůli tomu, že se tyto dva senzory výborně doplňují.

Rozpoznávání objektů ve scéně je možné využít v mnoha oblastech. Dá se použít pro řízení automobilů nebo robotů, zdokonalení využití energií v domácnostech a budovách obecně, záchranu lidí v nepřehledných a nebezpečných situacích a mnoha dalších odvětvích. Systémy pro rozpoznávání objektů mohou tedy zachraňovat životy. Automobil se vyhne překážce v případě nepozornosti řidiče, hasiči v hořící budově najdou uvězněného člověka v nouzi apod. Tyto systémy však mohou také pomáhat životnímu prostředí a to třeba regulací vytápění místností, ovládáním osvětlení podle toho kdo a kde se nachází a podobně.

Cílem této bakalářské práce je pak využít schopností milimetrového radaru a kamery pro rozpoznání určitých objektů. Je nutné zvolit způsob jakým budou tyto senzory spolupracovat. Dále je třeba navrhnout vhodnou aplikaci, kde je možné tento způsob rozpoznávání objektů použít a ověřit funkčnost řešení.

Následující kapitola pojednává o současném stavu metod a technologií, které souvisí s problematikou fúze milimetrového radaru a kamery. V kapitole **Zhodnocení aktuálního stavu a návrh řešení** jsou pak současné způsoby fúze a detekce zhodnoceny a je navrženo řešení, které bude využito. V kapitole **Popis vlastní práce** se nachází popis způsobu řešení daného zadání a testování výsledného systému.

Kapitola 2

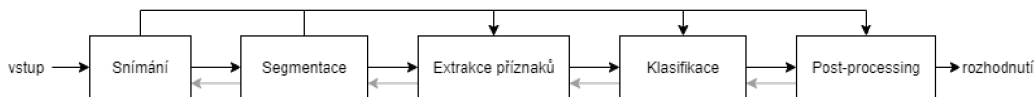
Shrnutí dosavadního stavu

Tato kapitola se věnuje seznámení čtenáře s termíny, technologiemi a metodami, které je možné použít při návrhu, realizaci a popisu vlastního řešení. Jelikož jsou použité senzory důležitou součástí této práce, čtenář bude seznámen i se základními principy fungování a snímání milimetrového radaru.

2.1 Rozpoznávání objektů

Pojem rozpoznávání se dá vysvětlit jako postup pro zařazování objektů (jevů a situací) reálného světa do tříd. Každý objekt, jev nebo situace jsou při dostatečném množství informací jedinečné [25], aby tedy bylo možné rozpoznat jistou skupinu objektů, je na ně potřeba nahlížet z určitého úhlu pohledu. Je třeba stanovit hledisko, podle kterého budou objekty rozděleny do skupin (tříd). To znamená, že je třeba určit veličiny, které určují do jaké třídy objekt spadá, a u těchto veličin je nutné určit s jakou frekvencí a přesností budou shromažďována. Toto se obecně nazývá „definice systému“ [25], problematice snímání se dále věnují části **Možnosti snímání objektů pomocí kamery** a **Možnosti snímání objektů za pomoci milimetrového radaru**. Výstupem snímání jsou pak data, která slouží jako vstup pro rozpoznávání. Je přitom vhodné minimalizovat množství zpracovaných dat a maximalizovat diskriminační schopnost (přednostně jsou vybrána data, na kterých při rozhodování nejvíce záleží), více viz sekce **Extrakce a výběr příznaků z dat kamery a milimetrového radaru**. Další částí, kterou je možné do zpracování zařadit je segmentace. Je výjimkou mít pro rozpoznávání k dispozici segmentovaná data, a proto je v některých případech vhodné tento krok zahrnout, více viz sekce **Segmentace objektů za pomoci dat z kamery**. Dále je již možné vzorky klasifikovat do jednotlivých tříd a to různými způsoby, tato práce se věnuje především použití konvolučních neuronových sítí, více viz sekce **Klasifikace objektů**. Závěrem je možné do rozpoznávání zanést kontextuální informace – post-processing [25].

Rozpoznávání a jemu předcházející zpracování vstupních informací je tedy možné rozdělit na několik částí. Ty jsou vyobrazeny na schématu 2.1, které je mírně pozměněnou reprezentací typického postupu rozpoznávání [19]. Přehledně je zde vyjádřeno pořadí jednotlivých operací, které jsou naznačeny šipkami, jenž také napovídají tomu, že většinu kroků lze vypustit. Vyobrazené části jsou dále popsány v následujících kapitolách. Důraz je kladen na spojitost s použitými technologiemi a senzory.

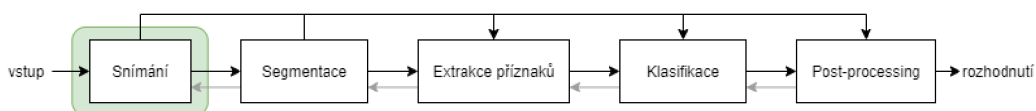


Obrázek 2.1: Postup rozpoznávání

2.2 Možnosti snímání objektů pomocí kamery

Snímání je pro dosažený výsledek velmi významné. Jedná se o stanovení veličin, které lze o rozpoznávaných předmětech zjistit, a způsob, jakým mohou být detekovány. Může to být váha, velikost, teplota, obraz z kamery, mračno bodů vytvořené senzorem (radar, lidar) atd. Je zapotřebí zohlednit cenu, přesnost jednotlivých způsobů získání dat a mnoho dalších parametrů [19].

Tato část se zaměřuje především na sběr dat z monokulární kamery. Klade důraz na rozlišení rozpoznávaného obrazu potřebného pro identifikaci objektů. Snímání je prvním krokem na cestě 2.2 k zařazení vstupu do třídy.



Obrázek 2.2: Snímání v kontextu procesu rozpoznávání

Snímání se nezahrnuje do samotné úlohy rozpoznávání a neexistuje žádná obecná metoda pro stanovení parametrů snímání. Je třeba určit jaké veličiny a s jakou frekvencí budou sledovány na základě tříd a případu/případů užití [25]. Pro účely detekce překážek jedoucího automobilu je tedy třeba data sbírat a vyhodnocovat často [23]. Naopak pro rozpoznávání počtu osob čekajících ve frontě je možné data sbírat s nižší frekvencí, jelikož v pozorovaném systému nedochází k častým změnám, a není tedy třeba na jeho změnu reagovat co nejrychleji.

Dalším základním kritériem je přesnost sběru dat. U obrazových dat je touto přesností míněno hlavně rozlišení. Studie ukazují, že redundance v obrazových datech může dosahovat až 99 % [27]. Díky tomu je možné snížit rozlišení, tím počet pixelů, a tedy i množství dat, které je nutno zpracovat, a to bez ztráty informací. Ve studii zabývající se minimálním počtem pixelů pro rozpoznání různých druhů objektů [27] bylo zjištěno, že člověk potřebuje pro rozpoznání lidské postavy obrázky o velikosti pouze 32×32 pixelů, což tvrzení podporuje.

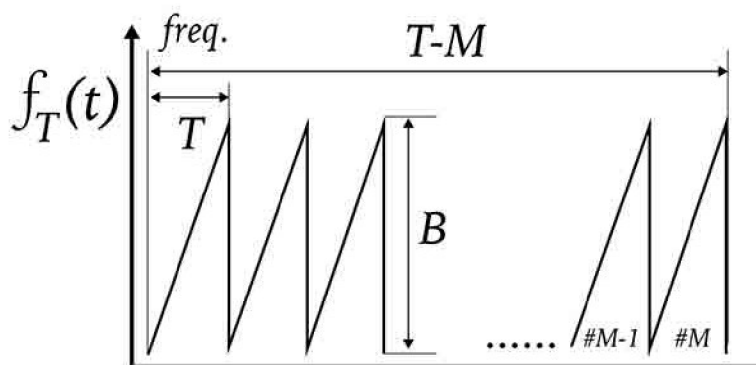
V dnešní době se ale při rozpoznávání objektů na obrazových datech k redukcí rozlišení příliš často nepřistupuje, a to především kvůli rozmachu konvolučních neuronových sítí a navýšení výkonu výpočetní techniky. Lze tedy s jistotou tvrdit, že neztrácíme žádné informace. V konvolučních neuronových sítích se v dnešní době používají obrázky s rozlišením 112×112 pixelů, 224×224 pixelů i 300×300 pixelů. Některé metody, označované jako jednokrokové, dokonce využívají obrázky 448×448 , více viz část 2.7.

U rozpoznávání je ale nejdůležitější, jaké veličiny jsou sledovány, podle jakých příznaků lze objekt nejlépe rozeznat. Pokud je nutné od sebe rozlišit dva objekty, které mají často velmi podobnou váhu, tato veličina by neměla být použita. Tímto úkolem se zabývá část **Extrakce a výběr příznaků z dat kamery a milimetrového radaru**. Následující podkapitola pojednává o způsobu jakým pracuje v této práci využívaný senzor – milimetrový radar 2.3.

2.3 Princip fungování a způsob snímání milimetrového radaru

Milimetrový radar, také označovaný jako mmWave radar, je typ senzoru pracující na základě elektromagnetického vlnění. Vyslaný signál se odrazí od objektů a vrací se k radaru, který vlnění přijímá, a tak objekty detekuje. V tomto ohledu se podobá svým předchůdcům, liší se od nich však délkou vysílaných vln, které mají vlnovou délku, uhádli jste, v řádech milimetrů. Tím je možné celý systém miniaturizovat, antény mohou být, stejně jako vysílané vlny, velmi malé [18].

Milimetrové radary však většinou nevysílají signál o konstantní frekvenci, místo toho šíří frekvenčně modulovaný spojitý signál (FMCW), který je znázorněn na obrázku 2.3. Jedna perioda (na obrázku 2.3 označena T) začínající frekvencí f_{min} a končící frekvencí f_{max} bude dále označována anglickým názvem „chirp“. Pomocí FMCW radaru je možné měřit vzdálenost, rychlost a úhel [18].



Obrázek 2.3: Způsob vysílání FMCW signálu [5]

Vzdálenost objektu lze zjistit jednoduše. Stačí měřit čas mezi vysláním signálu a jeho návratem. Tento čas poté můžeme díky rychlosti šíření tzn. rychlosti světla vypočítat. Pokud je několik objektů v zákrytu, lze je rozeznat díky Fourierově transformaci (FT), v přijímaném signálu se totiž objeví více složek frekvencí, které FT dokáže separovat [18].

Aby bylo možné zjistit rychlost detekovaného bodu, je třeba měřit změny kratší, než je perioda signálu. Tohoto je možné docílit za pomoci změny fáze vracejícího se signálu oproti signálu vyslanému. Za tímto účelem radar vyšle rychle po sobě dva chirpy (dvakrát signál začínající na frekvenci f_{min} a končící na f_{max}) o totožných frekvencích. Pokud se objekt pohybuje, bude mít přijímaný signál jinou fázi a je tedy možné z této změny vypočítat rychlost pohybu objektu. Změnu fáze signálu lze zjistit pomocí Fourierovy transformace. Maximální detekovatelná rychlost je podmíněna časem mezi vyslanými chirpy [18].

Pro nalezení úhlu, pod kterým je objekt detekován, jsou třeba dvě přijímací antény. Úhel pak lze zjistit na základě rozdílu vzdálenosti antén od objektu. K jedné z antén bude totiž signál putovat déle, a tím budou mít přijímané signály jinou fázi, kterou zjistíme opět aplikací Fourierovy transformace. Jedná se o podobný způsob jako u zjišťování rychlosti detekovaného objektu [18].

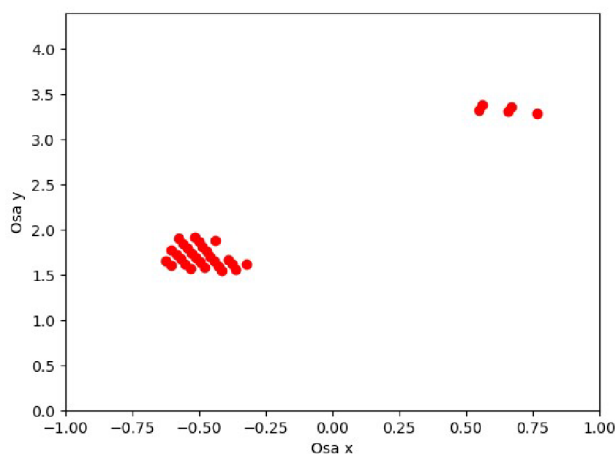
Milimetrový radar je tedy skvělý prostředek pro detekování objektů v prostoru. Díky malé vlnové délce je přesný a dokáže detekovat objekty, které jsou v zákrytu. Jaká data lze z mmWave radaru získat? Na to odpovídá následující podkapitola.

2.4 Možnosti snímání objektů za pomoci milimetrového radaru

Milimetrový radar je možné nastavit pro různé případy užití. Radar typu mmWave lze využít pro detekci osob v místnosti, monitorování provozu, detekci velmi malých vibrací atd. Tyto možnosti jsou zde zmíněny, jelikož není možné očekávat stejnou přesnost detekce při použití na vzdálenosti v řádu desítek metrů a v malé místnosti, toho není zmíněná technologie schopná.

Jak ale vlastně vypadají data, která milimetrový radar vytváří?

Je to mračno bodů (point cloud), které reprezentuje detekované objekty v prostoru. U těchto bodů radar detekuje vzdálenost, úhel, hodnotu SNR¹ a rychlost. Příkladem může být obrázek 2.4, kde pozice (0 0) reprezentuje pozici radaru a ostatní body – detekované objekty. Jelikož radar poskytuje pouze úhlové souřadnice, je pro toto zobrazení třeba převést body do kartézského souřadného systému, více viz část 2.8. Na obrázku 2.4 nelze vidět hodnoty SNR a rychlosti, každý z bodů těmito parametry disponuje.



Obrázek 2.4: Reprezentace mračna bodů v prostoru před radarem

Pro tuto práci byl použit milimetrový radar od společnosti Texas Instruments. Následující část se kvůli tomu zaměřuje na demonstrační příklady dodávané společně s radarem od tohoto výrobce a čerpá informace o možnostech snímání radaru z přednastavených konfigurací. Tyto demonstrační příklady často poskytují i další informace, jako jsou příslušnost bodů k různým shlukům, potažmo objektům a informace o těchto detekovaných objektech. Nemusíme tedy v některých případech hledat příslušnost jednotlivých bodů k různým předmětům v prostoru, ale jednoduše se spolehnout na algoritmy řešící tento problém již v radaru.

V případě již zmíněné aplikace – sledování provozu je možné pracovat s přesností 0,23 m pro vzdálenost objektu od radaru a 0,48 m pro rozlišení objektů nacházejících se vedle

¹SNR je zkratkou anglického Signal-to-Noise-Ratio a reprezentuje rozdíl mezi přijatým signálem a nastavenou maximální úrovní šumu. Zde reprezentuje něco jako jistotu, s jakou radar určitý bod detekoval [1]

sebe [4]. Maximální použitelná vzdálenost je pak až 185 m [4]. Tyto parametry snímání jsou závislé na nastavení vysílaného signálu. Rozlišitelnost objektů v rámci vzdálenosti od radaru závisí na délce trvání jednoho chirpu. Maximální vzdálenost je pak určena frekvencí vysílání jednotlivých chirpů. To tedy znamená, že se parametry ovlivňující rozlišitelnost objektů za sebou a maximální detekovatelnou vzdálenost vzájemně vylučují. Čím větší délka trvání chirpu (rozlišitelnost), tím menší frekvence chirpů (maximální vzdálenost) a naopak. Je tedy třeba vybrat konfiguraci na základě případu užití. Pro demonstrační příklad detekce osob v místnosti je výhodné nastavit mnohem menší maximální vzdálenost a získat tak větší rozlišení. V ukázkovém řešení od firmy Texas Instruments se jedná o rozlišitelnost 0,055 m [3].

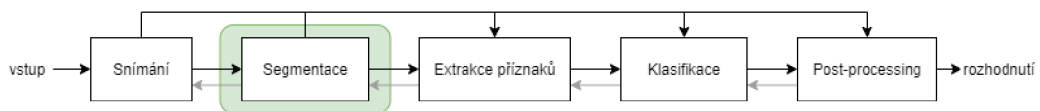
Poslední vlastností, kterou se bude tato část textu věnovat, je maximální detekovatelná rychlost. Ta je také závislá na periodě odesílání jednotlivých chirpů, musí tak být nastavena s ohledem na ostatní vlastnosti měření. V případě detekce automobilů je třeba detekovat objekty s vysokou rychlostí, demonstrační příklad podporuje rychlost až 65 km/h [4].

V tuto chvíli by měl být čtenář seznámen se základními principy fungování milimetrového radaru a možnostmi získávání dat z monokulární kamery a mmWave radaru. Problematikou extrakce a výběru příznaků z uvedených senzorů se pak zabývá část 2.6.

2.5 Segmentace objektů za pomoci dat z kamery

Segmentace je proces rozdělení vstupního obrazu na části podle společných vlastností [15] a může v procesu rozpoznávání předcházet klasifikaci 2.5. Existuje několik typů metod pro segmentaci obrazových dat:

- vycházející z detekce hran,
- orientované na regiony v obraze,
- založené na teorii grafů,



Obrázek 2.5: Segmentace v kontextu procesu rozpoznávání

Metody vycházející z detekce hran

Tento typ segmentace se snaží rozlišit hrany v obraze. Ty jsou detekovány pomocí hranových detektorů na základě rozdílu hodnot okolních pixelů. Tato metoda se právě kvůli tomuto způsobu detekce stává citlivou na šum. Výsledkem rozpoznávání je množina hran nalezených v obraze [15].

Metoda detekování hran lze použít pro řešení některých úkolů jako jsou: detekce řádků ručně psaného textu [8] nebo detekce objektů na monotónním pozadí. Ve většině případů však není vhodná, jelikož obraz na vstupu je příliš členitý.

Metody orientované na regiony v obraze

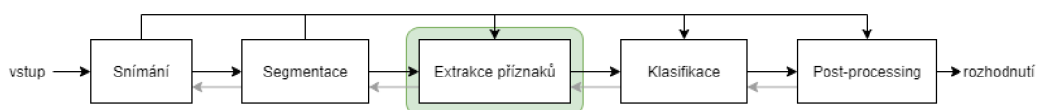
Metody zaměřující se na regiony v obraze oproti hranovým algoritmům detekují objekty na základě homogenity oblastí. Kritéria, která sledují, mohou být: úroveň šedi, barva, textura, tvar apod. Výhoda těchto metod spočívá v menší náchylnosti na chyby při zašumělém obraze [15].

Segmentace založená na teorii grafů

Segmentace využívající teorii grafů typicky využívá neorientovaný graf, který reprezentuje obraz na vstupu. Snaha je pak o rozdělení tohoto grafu na podgrafy, které reprezentují jednotlivé objekty. V praxi využívaným příkladem algoritmu je *Efficient Graph-Based Image Segmentation* [17].

2.6 Extrakce a výběr příznaků z dat kamery a milimetrového radaru

V praxi často není možné použít všechna data, která mohou být o objektu shromážděna pro klasifikaci. Máme příliš mnoho atributů, některá data nejsou pro klasifikaci důležitá apod. V procesu rozpoznávání se výběr a extrakce příznaků² typicky nachází před samotnou klasifikací 2.6.



Obrázek 2.6: Extrakce příznaků v kontextu procesu rozpoznávání

Jako příznaky použité pro rozpoznávání je možné zvolit již existující atributy, které již máme k dispozici (výběr příznaků), nebo je možné na základě získaných dat vytvořit nové příznaky, které budou pro úkol výhodnější (extrakce příznaků).

Výběr příznaků z dat radaru

Výběr příznaků spočívá v hledání množiny příznaků, která maximalizuje schopnost učícího algoritmu správně klasifikovat zpracovávaná data [10].

Jak je zmíněno v předchozí kapitole **Možnosti snímání objektů za pomoci milimetrového radaru**, data z radaru obsahují informace o jednotlivých bodech, jejich umístění, rychlost a jistotu s jakou byly detekovány, dále pak lze získat další informace o shlucích bodů, jako jsou příslušnost bodů k jednotlivým shlukům, souřadnice potenciálně detekovaného objektu apod.

Některé práce se tedy zaměřují a spoléhají se přímo na data týkající se shluků, s mračenem bodů nepracují, a zabývají se pouze tímto radarem detekovaným objektem. Mezi tyto práce patří systém pro detekci objektů (překážek) na základě dat z kamery a milimetrového radaru pro automobily [23]. Systém využívá jak pro kalibraci, tak pro samotnou detekci překážek právě radarem detekované shluky a tím algoritmus k-nejbližších sousedů, pomocí kterého detekují potenciální překážky automobilu.

Dalším možným přístupem je pro úlohu detekce (rozpoznávání) použití mračna bodů. Tento způsob nám může dát větší kontrolu a více možností pro shlukování a detekci. Tato data je možné použít jako vstup pro neuronové sítě a jiné výpočetně náročnější metody. Tento přístup využívá třeba práce [20] zabývající se zvýšením spolehlivosti detekce a sledování objektů za pomoci fúze dat z kamery a mmWave radaru.

Je samozřejmě možné použít všechna data, které nám jsou radarem poskytnuta. Rozšířit množství dat, která jsou vázána na jednotlivé body, o index shluku, ke kterému patří apod.

Extrakce příznaků z dat kamery

Extrakce příznaků spočívá v hledání zobrazovací množiny F do množiny F' , která maximalizuje schopnost učícího algoritmu správně klasifikovat zpracovávaná data. To znamená,

²Příznak je měřitelný atribut nebo charakteristika pozorovaného fenoménu (objektu, jevu, situace apod.) [19]. Příkladem příznaku může být: váha, barva, velikost atd.

že pokud nejsou k dispozici vyhovující příznaky, nebo jsou nedostatečné, je třeba vytvořit nové (za pomoci dostupných dat), které dokáží rozlišit třídy lépe než původní příznaky [10].

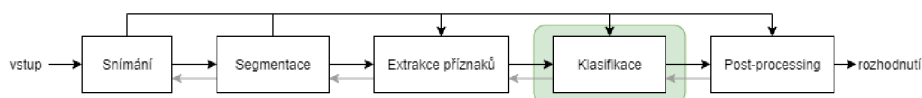
Z obrazových dat je možné získat mnoho příznaků, není tedy třeba se spoléhat pouze na hodnoty jednotlivých pixelů. Patří mezi ně:

- texturní příznaky,
- barvy (barevný histogram),
- rohy a hrany,
- tvar objektů a další [24].

Texturní příznaky mohou pomoci při vyhledávání určitých textur v obraze a při rozdělení na části [24]. Barva je pak jeden z velmi důležitých příznaků obrazu. Jako příznak je barvu možné využít ve formě barevných histogramů, které lze porovnávat oproti vzoru [24]. Rohy a hrany mohou být detekovány mnoha různými způsoby, mohou to být např. techniky založené na konvoluci, Houghova transformace [6] atd. Většina těchto příznaků se však často nepoužívá a má své určité aplikace. Extrakce příznaků z obrazových dat je velmi významná pro fungování konvolučních neuronových sítí, proč, je rozebráno v části 2.7.

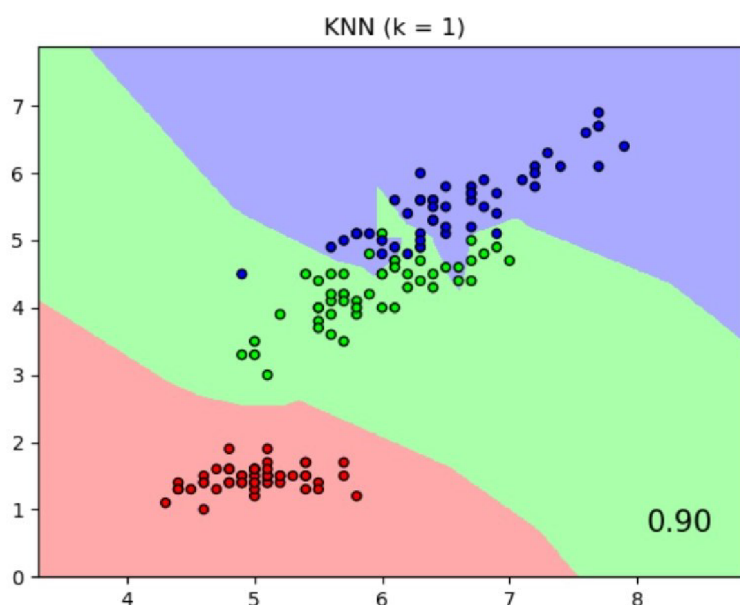
2.7 Klasifikace objektů

Tato část pojednává o další z částí procesu rozpoznávání – klasifikaci. Čtenář bude seznámen s obecnými principy klasifikace, fungováním neuronových sítí se zaměřením na konvoluční neuronové sítě (CNN) a s tím související problematikou extrakce příznaků. Pozornost bude věnována jednotlivým vrstvám CNN a architektuře konvoluční neuronové sítě pro účel fúze a klasifikace.



Obrázek 2.7: Klasifikace objektů v kontextu procesu rozpoznávání

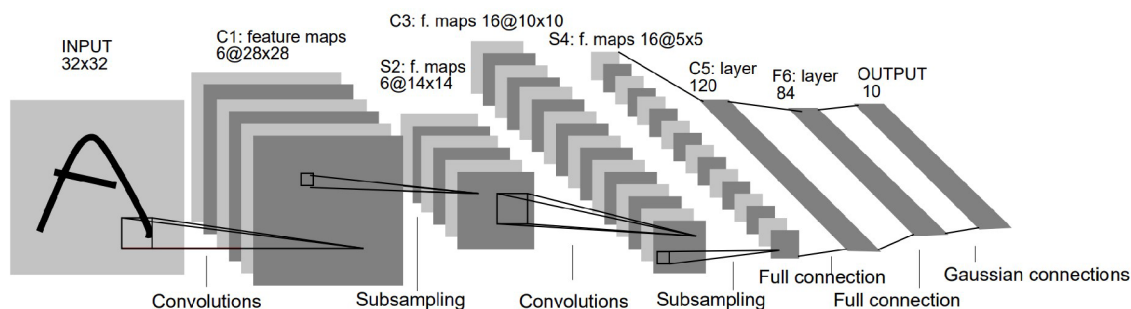
Úloha klasifikace je snaha o zařazení rozpoznávaného objektu do předem daných skupin (tříd), a to za použití předem připravených dat ke klasifikaci předcházejících kroků rozpoznávání [19]. K tomu jsou typicky třeba trénovací data, která použijeme pro nacvičení klasifikátoru. Tato cvičná data se objevují vždy ve dvojicích: množina příznaků a zařazení do třídy. Cvičení klasifikátoru pak slouží k nastavení „hranic“ jednotlivých tříd na základě hodnot příznaků jednotlivých vzorků. Je možné od sebe oddělit trénovací data, a tím objekty (pozorování) patřící do různých tříd. Příkladem nastavení hranic klasifikátoru může být obrázek 2.8. Jednotlivé hranice tříd jsou zde znázorněny jako přechody mezi barvami. Klasifikace pak spočívá v nalezení odpovídajícího bodu podle hodnot příznaků a zařazení objektu, který k tomuto bodu náleží, do příslušné třídy. V případě obrázku 2.8 jsou sledovány dva příznaky, jeden je zobrazen na ose x, druhý na ose y. Lze však pracovat s mnohem větším počtem příznaků, a tím i více dimenzemi.



Obrázek 2.8: Ukázka rozhodovacích hranic tříd [2]

Klasifikace objektů pomocí konvoluční neuronové sítě

Konvoluční neuronové sítě (CNN) jsou umělé neuronové sítě inspirované způsobem vidění zvířat [28]. Jsou složeny z vrstev: konvoluční, max-pooling a klasifikační. Vrstva sloužící pro klasifikaci je tvořena umělými neurony, které jsou v podstatě jednoduché klasifikátory - perceptrony. Konvoluční a pooling vrstvy pak slouží k nalezení příznaků v rámci neuronové sítě [7]. Oproti jiným typům umělých neuronových sítí, dokáží CNN pracovat s 2D kontextem, což je u rozpoznávání obrazových dat velmi důležité, a proto se staly oblíbeným prostředkem počítačového vidění [13]. Příkladem návrhu konvoluční neuronové sítě může být převzatý obrázek 2.9. Jedná se o grafické znázornění procesu rozpoznávání konvoluční neuronové sítě, které bylo navrženo profesorem Yann LeCunem v roce 1989 a dalo základ této metodě rozpoznávání [26]. Prvním krokem CNN je preprocessing dat a normalizace. Poté jsou extrahovány příznaky a redukována redundance pomocí konvolučních a pooling (sub-sampling) vrstev. Závěrem jsou všechny získané příznaky klasifikovány pomocí plně propojené vrstvy (případně vrstev), která/é poskytne ocenění pro každou rozpoznávanou třídu [13].



Obrázek 2.9: Typická posloupnost vrstev konvoluční neuronové sítě [28]

Konvoluční vrstva CNN

Konvoluční vrstva CNN je v praxi aplikace filtru za účelem extrakce příznaků z obrázku. První vrstvy většinou v obrazových datech hledají hrany, další se mohou zaměřovat na tvary apod.

Proč je ale vlastně třeba konvoluci na rozpoznávaná data aplikovat?

Konvoluce pomáhá hned v několika ohledech. Jedná se o extrakci příznaků, která zachovává 2D kontext obrázku, a hlavně redukuje počet parametrů neuronové sítě. Mějme obrázek o velikosti $28 \times 28 \times 3$ pixelů, pokud si toto převedeme na počet pixelů, dostaneme 2352 pixelů, v klasické umělé neuronové síti by toto znamenalo 2352 parametrů pro učení, což je už i pro tak malý obrázek mnoho. Dalším problémem je škálovatelnost, pokud by byl totiž použit větší obrázek o velikost třeba $300 \times 300 \times 3$ pixelů, což je 270 000 pixelů. Počet parametrů by byl již opravdu neúnosný, a to pouze v rámci jedné vrstvy. Konvoluční vrstvy tento problém dokáží řešit pomocí aplikace „filtru“. Dochází zde totiž ke sdílení vah mezi jednotlivými sektory [7]. Místo toho, aby existovala váha pro každý pixel obrázku, existuje pouze určitý počet vah daný velikostí filtru. Vychází se zde z jednoduchého předpokladu:

Pokud dává smysl získat určitý příznak na jednom místě, dává to smysl i na jiném místě v obrázku [7]. Na všech místech v obrázku dává smysl určit, zda se zde nachází hrana nebo ne apod.

Každá konvoluční vrstva je definována pomocí několika parametrů: velikost vstupu, velikost jádra, hloubka (počet vrstev vstupu), padding³ a stride⁴. Padding zvětšuje výslednou velikost výstupu po aplikaci konvoluce, je tedy možné využít více konvolučních vrstev po sobě [7]. Padding také bojuje se ztrátou informací v okrajích obrázku. Krajní pixely obrázku totiž budou pro vstup filtru použity pouze jednou, zatímco středové hned několikrát. Stride oproti padding naopak snižuje velikost výsledku konvoluce.

Pooling vrstva CNN

Pooling vrstva se používá k redukcí dimensionalit. Existuje několik strategií pro realizaci pooling vrstvy: max-pooling, average-pooling a další [7]. Nejpoužívanější je přístup max-poolingu, kdy se z dané oblasti vybere největší hodnota, která se pak stane výstupem.

Aktivační funkce CNN

Výstupy každé z vrstev mohou nabývat hodnot od $-\infty$ až po $+\infty$. Funkce produkující výstupy není nikterak omezená:

$$Y = \sum (weight \times input) + bias \quad (2.1)$$

Je potřeba určit aktivační funkci, která zajistí, že výstupem perceptronu budou hodnoty omezené, nejčastěji mezi nulou a jedničkou. Reálné neurony se totiž chovají podobně, buď jsou neaktivní, nebo plně aktivní. K tomuto účelu se používají následující funkce: schodová, sigmoida, tanh, ReLu a další [7].

Každá ze zmíněných funkcí má své výhody a nevýhody. Schodová funkce je nejjednodušší způsob omezení vstupu – jednoduše získáme hodnotu „pravda“ nebo „nepravda“. Problém s tímto řešením nastává, pokud je třeba klasifikovat více tříd. V tu chvíli může mít hodnotu „pravda“ více perceptronů a není možné rozhodnout, do které třídy objekt zařadit. Tento popsaný problém řeší ostatní zmíněné metody, které mohou nabývat hodnot v určitém intervalu, a nefungují tedy pouze binárně. Co se týče výběru mezi těmito spojitými metodami, neexistuje obecně nejlepší aktivační funkce, každá se hodí na něco jiného. Pokud jsou známy vlastnosti aproximované funkce, je podle toho možné vybrat aktivační funkci. Nejjednodušší učící proces by měla mít ve většině případů metoda ReLu, je s ní tedy dobré začít a až poté zkoušet další, nebo vytvářet vlastní aktivační funkce [7].

Klasifikační vrstva CNN

Klasifikační vrstva je zpravidla plně propojená, nedochází tedy ke sdílení parametrů jako u konvoluční vrstvy. Platí, že jedna hodnota na výstupu odpovídá ohodnocení jedné konkrétní třídy. Výstupem je tedy vektor hodnot odpovídající ohodnocením jednotlivých tříd. Klasifikace pak probíhá jednoduše výběrem třídy s nejvyšším ohodnocením. Pokud je více tříd ohodnoceno vysokou známkou, je možné opět vybrat třídu s nejvyšším ohodnocením, nebo jiným způsobem určit výsledek zahrnutím kontextuálních informací.

³Padding je přidání okrajů s nulovou hodnotou k obrázku.

⁴Pojmem stride se rozumí krok, o který bude pokaždé aplikovaný filtr posunut.

Architektura konvolučních neuronových sítí

Tato část se zabývá tím, jak vytvořit z jednotlivých vrstev fungující konvoluční neuronovou síť. Jak je zmíněno výše, konvoluční neuronová síť se typicky skládá ze tří typů vrstev (konvoluční – CONV, pooling – POOL a plně propojená – FC). Jako vrstva bude v této práci vyobrazována i aktivační funkce (např. ReLu).

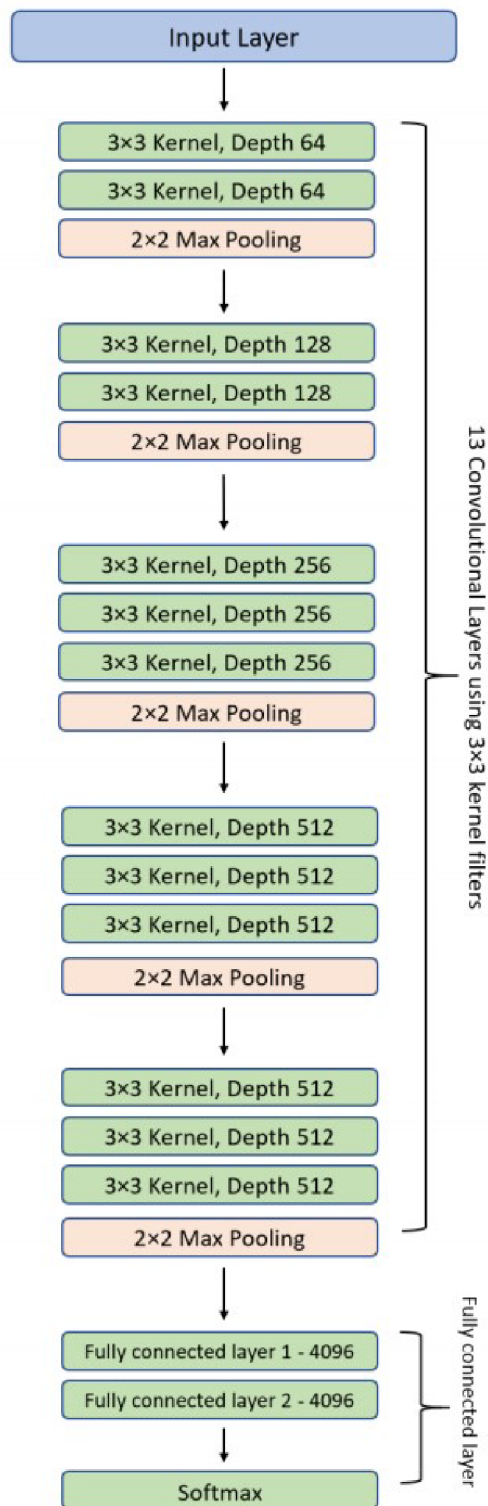
Nejběžnějším způsobem skladby vrstev je využití několika kombinací CONV–RELU vrstev, která jsou následovány POOL vrstvou. Tato kombinace je běžně opakována do té doby, dokud nejsou dimenze obrázku velmi malé. Poté se využijí plně propojené vrstvy, poslední dává výstup klasifikace [7]. Výše popsany obecný návrh může být reprezentován následujícím vzorem [7]:

$$INPUT \rightarrow [[CONV \rightarrow RELU] * N \rightarrow POOL?] \rightarrow [FC \rightarrow RELU] * K \rightarrow FC \quad (2.2)$$

kde „*“ značí opakování a „?“ nepovinnou vrstvou. N většinou nabývá hodnot menších, nebo rovno třem. Počet plně propojených vrstev je typicky menší než tři ($K < 3$). Pokud se má jednat o CNN, pak musí platit: $M > 0$.

Obecně platí, že by měl návrh pro CONV vrstvy používat více menších CONV jader (filtrů). Nejčastěji se používají jádra o velikosti 3×3 . Je to opět kvůli počtu parametrů, které je třeba nastavit a uložit. Dále pak platí, že jádra o velikosti 3×3 vyhledávají lepší, zajímavější příznaky. Co se týče POOL vrstev, zde opět platí: čím menší, tím lepší. Obvykle se používají jádra o velikosti 2×2 a $Stride = 2$. Literatura [7] také radí, že by se často k návrhu konvoluční sítě vůbec nemělo přistupovat a spíše použít řešení, které v současnosti nejlépe funguje na datasetu ImageNet a dotrénovat CNN na svých trénovacích datech. Ne vždy je však tento přístup možné použít.

Na závěr ještě ukázka architektury CNN VGGNet [22], na obrázku 2.10 je znázorněna posloupnost aplikace jednotlivých typů vrstev na vstup. Lze si také povšimnout, že se přesně drží výše zmíněných doporučení. Konvoluční vrstvy používají velikost jádra 3×3 a jelikož ze vstupu $224 \times 224 \times 3$ získává data o stejných rozměrech, lze vyvodit, že: $Stride = 1$ a $Padding = 1$. Dále se pak rozvržení vrstev řídí obecným vzorem 2.2.



Obrázek 2.10: Architektura konvoluční neuronové sítě VGGNet [22]

Proces učení neuronových sítí

Proces učení je u každé neuronové sítě (NN) velmi důležitý. Je třeba najít co nejlepší hodnoty vah perceptronů neuronové sítě, tak, aby vyhovovaly daným požadavkům. K učení je třeba mít trénovací data, která se skládají ze souboru příznaků a zařazení do určité třídy („label“), jedná se tedy o učení s učitelem. Implementuje se jako iterativní proces, který se dá rozdělit na několik částí: forward-propagation, loss function a back-propagation [7].

Forward-propagation je pouze vyhodnocení trénovacích dat pomocí neuronové sítě a shromáždění výsledků. Ke každému vstupu NN předpoví třídu, která bude dále porovnána se známou skutečnou třídou [7].

Následuje vyhodnocení toho, jak dobře NN pracuje. Vypočte se pomocí ztrátové funkce, která porovná předpovězené výsledky a skutečné třídy, kterým data odpovídají. Výsledkem je pak číslo značící jak dobře NN přiřazuje třídy k trénovacím datům (jak klasifikuje) [7].

Nyní je třeba na základě výsledku ztrátové funkce upravit váhy jednotlivých perceptronů. Tento úkol plní back-propagation. Propaguje výsledek zpět skrz neuronovou síť, která musí zareagovat úpravou vah. Ke každému neuronu (perceptronu), se ale musí dostat jiná informace, jelikož se každý neuron na výsledku klasifikace podílel jinak [7].

Nejpoužívanějším způsobem optimalizace vah je algoritmus klesání podle gradientu (gradient descent), který vychází z první derivace ztrátové funkce (gradientu). Gradient je vektor, jehož směr je shodný se směrem maximálního růstu funkce. To tedy znamená, že pokud použijeme pro úpravu vah zápornou hodnotu gradientu, budeme minimalizovat hodnotu ztrátové funkce. Gradient je třeba vypočítat pro každý z neuronů a je třeba zohlednit jejich aktivační funkce. Algoritmus klesání podle gradientu je možné implementovat ručně za pomoci chain rule, nebo lze využít optimalizátor, který hodnoty gradientů vypočítá. Tomu stačí zadat hodnotu, o kterou má váhy upravovat – míru učení (learning rate), která určuje, jak rychle NN reaguje na změny, a tím získat jednotlivé sady gradientů [7].

Výše popsané kroky se opakují, dokud ztrátová funkce není dostatečně nízká, nebo není vykonán dostatečný počet iterací.

Způsoby hledání oblastí zájmu pro klasifikaci

Aby bylo možné v obraze detekovat objekt a ne jej jen klasifikovat, je třeba klasifikátoru (neuronové sítě apod.) poskytnout jednotlivé části obrazu, které jsou pak klasifikovány a v obraze tak nalezeny jednotlivé objekty.

Problém je, že není možné předem určit, jak velké objekty je třeba hledat, a kde se v obraze nacházejí. Je tedy třeba vyzkoušet mnoho oblastí. Každá oblast je při tom poslána do klasifikátoru, který zjistí, jestli se jedná o vyhledávaný objekt, či nikoliv. Tím je tento přístup velmi náročný na výkon a čas, proto se v praxi často nevyužívá. Dá se použít pro detekování jedné třídy objektů pro daný poměr stran [29].

Výše popsaný přístup se nazývá metodou *Sliding Window* – metoda okénka a je tím nejintuitivnějším přístupem k řešení tohoto problému. Je to však také klasický případ řešení pomocí hrubé síly. Navrhovaných oblastí může být řádově i několik desítek tisíc. Není možné počet navrhovaných okének snížit?

Mezi v dnešní době využívané metody pro efektivnější výběr okének patří *selective search* [12]. Ta využívá výše zmíněný postup Pedra Felzenszwalba a Daniela Huttenlochera *Efficient Graph-Based Image Segmentation*. Dochází při tom k iterativnímu seskupování segmentů obrazu. Algoritmus začíná na úrovni pixelů a postupně části seskupuje. Zároveň nalezené segmenty uchovává a přidává do seznamu potencionálních objektů. Postup vypadá následovně:

1. seskup segmenty podle jejich podobnosti,
2. ulož nově vzniklé segmenty do seznamu potencionálních objektů,
3. přejdi na první krok, pokud existuje více než jeden segment,
4. vrať seznam nalezených oblastí.

Jako kritérium podobnosti regionů se používá barva, textura, velikost a tvar. Pro finální ohodnocení podobnosti je pak použita lineární kombinace vektorů.

$$s(r_i, r_j) = a_1 s_{color}(r_i, r_j) + a_2 s_{texture}(r_i, r_j) + a_3 s_{shape}(r_i, r_j) + a_4 s_{size}(r_i, r_j)$$

Kde r_i a r_j značí dva segmenty obrazu a $s(r_i, r_j)$ je výsledné skóre podobnosti těchto oblastí. Díky tomu, že je tato metoda iterativní, a že postupně vyhledává objekty s různým zvětšením, je pro segmentaci a následné rozpoznávání vhodná.

Metody, jako *selective search*, sice rychlost vyhledání okének (vymezené oblasti objektu) zvýší, stále se však jedná řádově o sekundy. Značné zrychlení nabízí začlenění vyhledávání okének do neuronové sítě. Právě o to se snaží jednopřechodové metody, mezi které patří YOLO („You Only Look Once“) nebo SSD („Single Shot Detection“). Tyto metody nezačínají zdola (od pixelů) a postupně neslučují jednotlivé části, ale shora (od vstupního obrazu). Výsledkem takových CNN jsou pak ohodnocené regiony obrázku, které značí jaká třída se v dané části nachází. Následující část se zaměřuje na způsob fúze dat z mmWave radaru a kamery [2.8](#).

2.8 Fúze dat kamery a milimetrového radaru

Tato část se věnuje jednotlivým způsobům fúze sensorických dat z milimetrového radaru a monokulární kamery. Popisuje dále možnosti nalezení transformace mezi body milimetrového radaru a kamery.

Prvním způsobem fúze je již zmíněný článek popisující řešení problému detekce překážek pro automobil za pomoci právě mmWave radaru a kamery [23]. Milimetrový radar je zde využíván pro hledání potencionálních sektorů, kde se může překážka nacházet. Jak ale tento detekovaný objekt, potažmo sektor, najít v souřadném systému kamery? Kamera a mmWave radar na scénu přece pohlíží jiným způsobem. Radar získává pohled shora, zatímco souřadný systém kamery je jakýsi průřez prostoru pod určitým úhlem. Je tedy třeba najít způsob, jakým lze body ze souřadného systému radaru přenést do souřadnic kamery, nebo naopak. V případě zmíněného článku je to případ první. Zmíněný problém řeší i studie snažící se o zvýšení spolehlivosti detekce mmWave radaru [20]. Ta pro zlepšení výsledků využívá informace o nedokonalostech kamery (radiální distorze způsobená nedokonalostmi čočky kamery apod.) a snaží se o jejich odstranění. Řídí se prací, která zkoumá vztah souřadného systému kamery a 3D prostoru [11]. Jak práce [23], tak práce [20] řeší uvedený problém pomocí metody nejmenších čtverců.

Transformace bodů – metoda nejmenších čtverců

Tato metoda transformace bodů ze souřadného systému radaru do systému souřadnic kamery vychází z předpokladu, že dokážeme objekty detekovat oběma senzory, musí tedy existovat určitý vztah mezi souřadnicemi stejného objektu pro kameru i radar. Musí tím existovat i transformace vyjadřující tento vztah:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = T \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} \quad (2.3)$$

kde u a v jsou souřadnice určitého bodu v souřadném systému kamery a x a y jsou souřadnice v systému souřadnic radaru převedené z úhlových souřadnic do kartézských díky vztahu:

$$(x, y) = (range \times \sin(angle), range \times \cos(angle)) \quad (2.4)$$

kdy „range“ je vzdálenost detekovaného objektu od radaru a „angle“ úhel, pod kterým byl detekován. Z-ové souřadnice kamery i radaru nejsou k dispozici. T je pak transformační matice (3×3), která reprezentuje transformaci mezi souřadnicemi radaru a kamery. Transformační matice může být rozepsána následovně:

$$T = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} t_{1i} \\ t_{2i} \\ t_{3i} \end{pmatrix} \quad (2.5)$$

Jednotlivé hodnoty matice T je možné získat pomocí metody nejmenších čtverců, a to za pomoci Moore-Penroseovi pseudoinverzní matice. Platí totiž:

$$t_{1i} = (XX^T)^{-1} X^T * U \quad (2.6)$$

$$t_{2i} = (XX^T)^{-1} X^T * V \quad (2.7)$$

$$t_{3i} = (XX^T)^{-1} X^T * I_{n \times 1} \quad (2.8)$$

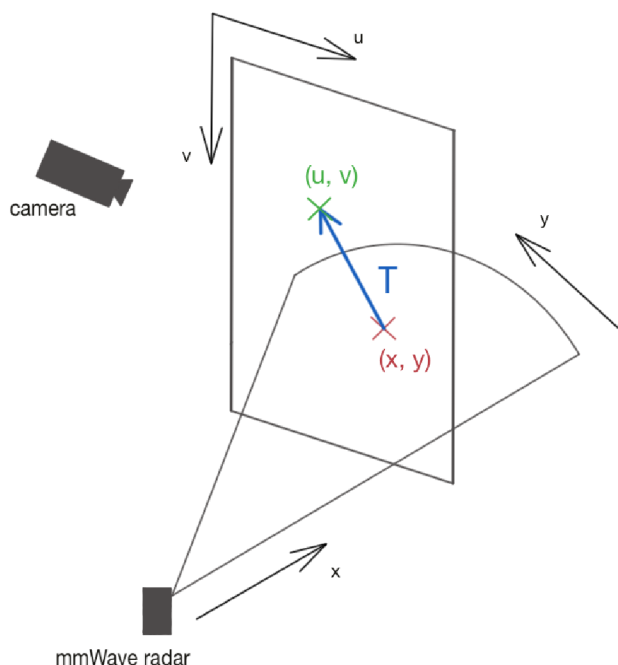
kde:

$$X = \begin{pmatrix} x_1, & y_1, & 1 \\ x_2, & y_2, & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{pmatrix} \quad (2.9)$$

$$U = (u_1, u_2, \dots u_n) \quad (2.10)$$

$$V = (v_1, v_2, \dots v_n) \quad (2.11)$$

a I je jednotková matice, n pak počet měření a $(XX^T)^{-1}X^T$ je pseudoinverzní matice realizující metodu nejmenších čtverců. Touto cestou je tedy možné provést transformaci ilustrovanou na obrázku 2.11. (x, y) jsou souřadnice radaru, (u, v) jsou souřadnice kamery a T je transformace odpovídajícího bodu radaru do souřadného systému kamery.



Obrázek 2.11: Transformace bodů ze souřadného systému radaru do systému kamery

Výše zmíněnému způsobu je zde věnována pozornost, jelikož byl původně použit pro vlastní řešení bakalářské práce. Ukázalo se však, že zjištěná transformační matice není pro účely projekce bodů do obrazu dostatečná, více viz sekce 4.3. Dále je popsán druhý možný postup pro kalibraci senzorů na základě informací o rozmístění senzorů v prostoru a jejich parametrech (vnější a vnitřní parametry), pracuje tedy s daty ve třech dimenzích.

Transformace bodů radaru pomocí vnitřních a vnějších parametrů

Při použití následující metody je třeba nejprve provést transformaci bodů ze souřadného systému mmWave radaru do světové soustavy souřadnic. Není však možné z 2D mračna

bodů zjistit výšku, v jaké se body nacházejí. Je tedy třeba tuto informaci kalibrační metodě dodat, nejsnazším způsobem je přiřazení informace o výšce bodů na základě umístění antén mmWave radaru. Tím je každý bod definován souřadnicemi (x_w, y_w, z_w) (x_w a y_w jsou zjištěny z bodů radaru na základě vztahu 2.4).

Transformace ze světového souřadného systému do soustavy souřadnic kamery je pak dána následujícím vztahem:

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (2.12)$$

kde R je matice (3×3) definující natočení a t je vektor určující pozici kamery. Jak matice R , tak vektor t jsou spjaté s prostorem světa, je tedy třeba určovat posunutí počátku souřadnic světa v systému souřadnic kamery a směry os soustavy souřadnic světa v systému kamery. Toto je neintuitivní způsob, který lze modifikovat. Je totiž možné popsat umístění a natočení kamery v prostoru světa a poté nalezenou matici invertovat.

Jakmile jsou k dispozici body v soustavě souřadnic kamery, je třeba už jen promítnout nalezené body do obrazové roviny. K tomuto účelu je třeba získat parametry kamery: ohniskové vzdálenosti f_x, f_y a hlavní snímkový bod (středový) P , se souřadnicemi p_x a p_y . Projekci do obrazu lze pak vyjádřit vztahy:

$$u = f_x \times \frac{x_w}{z_w} + p_x \quad (2.13)$$

$$v = f_y \times \frac{y_w}{z_w} + p_y \quad (2.14)$$

kde u a v jsou jako v případě první metody souřadnicemi v obrazové rovině. Tato kalibrace lze tedy uskutečnit za pomoci vnitřních parametrů (ohniskové vzdálenosti, hlavní středový bod) a vnějších parametrů (rozmístění senzorů v prostoru). Podobný postup využívá práce [9] zaměřující se na přesnou kalibraci mmWave radaru a kamery pro 3D rekonstrukci budov.

Klasifikace pomocí dat z kamery a milimetrového radaru

Výše zmíněné práce [23, 20] využívají pro kalibraci metodu první. Práce [23] zabývající se detekcí překážek v prostoru, pokračuje rozpoznáváním objektů nacházejících se v sektorech, kde milimetrový radar detekuje přítomnost objektu. K tomuto úkolu již však nevyužívá data zjištěná mmWave radarem. Z hlediska klasifikační metody tedy není pro účely této práce významná. V druhé práci [20] jsou použity algoritmy pro detekci pohybu a rozpoznávání nezávisle pro radar i kameru a posléze dochází k ověření výsledků mezi sebou. Data z radaru a kamery jsou tedy použity samostatně a nejsou vstupem jedné klasifikační metody. . .

Při hledání dostupných informací nebyl nalezen příklad využití dat z kamery i mmWave radaru jako vstupu jedné klasifikační metody, čehož se tato práce snaží docílit.

Pro fúzi senzorických dat je možné použít konvoluční neuronovou síť (CNN) a to s vynikajícími výsledky [14]. CNN také zajišťuje extrakci příznaků a klasifikaci samotnou (více viz sekce 2.7). Mělo by tedy být možné využít data z mmWave radaru i z kamery jako vstup pro CNN a zlepšit fungování systému za zhoršených světelných podmínek nebo zobrazit vzdálenost detekovaného objektu v obraze z kamery díky informacím z milimetrového radaru apod.

Kapitola 3

Zhodnocení aktuálního stavu a návrh řešení

V této části je popsáno zhodnocení metod, které jsou nyní používány pro fúzi senzorických dat z kamery a milimetrového radaru. Dále je pak navržen možný způsob řešení fúze dostupných senzorů za pomoci konvoluční neuronové sítě, který je ohodnocen pro různé případy k ověření navrhovaného systému. Na konci této kapitoly se nachází technické zadání pro následující část **Popis vlastní práce**.

3.1 Shrnutí dosavadního stavu fúze milimetrového radaru a kamery

Jak je popsáno v části 2.8, zatím není obvyklé aby se používal systém, který by pro rozpoznávání objektů využíval kameru a milimetrový radar. Existují studie[23, 20], které tyto senzory používají společně, v rámci této práce však nebyl nalezen případ, kdy by byla data z obou senzorů vstupem jedné klasifikační metody. Kamera je obecně výborným prostředkem pro detekci a rozpoznávání objektů, má však svá úskalí, mezi které patří špatné světelné podmínky, prach, mlha a jiné úkazy z principu bránící kameře ve správném fungování, které však nejsou překážkou (obrazně i doslovně) milimetrovému radaru. Existuje také způsob jak jednoduše provést fúzi dat z více zdrojů – neuronová síť [14].

Proč této situace nevyužít a nenavrhnout systém, který data z mmWave radaru a kamery využívá k detekci pomocí konvoluční neuronové sítě?

3.2 Návrh řešení fúze milimetrového radaru a kamery

Na základě vyhodnocení současného stavu jsem se tedy rozhodl systém, který tento způsob fúze využívá, navrhnout a vyzkoušet si tak technologické možnosti a vlastnosti tohoto spojení. Tato práce se tedy pokouší o využití dat z milimetrového radaru a kamery pro fúzi konvoluční neuronovou sítí. Pro tento účel jsem musel stanovit vhodný případ k ověření funkčnosti fúze dat senzorů, na základě kterého bylo nutné zvolit vhodný demonstrační příklad pro milimetrový radar. Dále jsem pak potřeboval získat data z radaru, které je nutné transformovat, upravit a zkalibrovat s kamerou. Pro účely kalibrace jsem musel navrhnout vhodný způsob umístění použitých senzorů v prostoru a stanovit postup kalibrace. Jelikož se navržené řešení nemohlo spoléhat pouze na obrazová data, nebylo možné použít pře-

dem natrénované konvoluční neuronové sítě (CNN). Bylo tedy třeba navrhnout vhodnou architekturu CNN, získat trénovací a testovací data, na kterých bude CNN trénována.

3.3 Vhodné případy k ověření fúze milimetrového radaru a kamery pro detekci

Jak je zmíněno výše, a detailněji řešeno v části 2.4, bylo třeba vybrat z řady vhodných případů užití, které mohou být použity pro ověření vlastností navrženého systému. Mezi ně jistě patří případy užití, které jsou pokryty demonstračními příklady pro milimetrové radary od firmy Texas Instruments. Některé z nich jsou uvedeny v části 2.4 a patří mezi ně počítání osob v místnosti, monitoring dopravy, detekce životních funkcí, rozpoznávání gest atd. Zmíněné případy užití by tedy měly díky fúzi s kamerou benefitovat a stát se spolehlivějšími. Další možností byl návrh vlastních případů k ověření vlastností systému, na které by kamera ani mmWave radar za určitých okolností samy nestačily.

Mezi tyto případy spadají: rozpoznávání osob ve zhoršených podmínkách viditelnosti nebo v zákrytu, zobrazování informací o detekovaných objektech do obrazu (vzdálenost, rychlost, směr pohybu), zvýšení spolehlivosti rozpoznávání pro kameru podobných objektů pomocí informací z mmWave radaru, detekce překážek automobilu a další.

V této práci jsem se rozhodl dále realizovat případ užití: rozpoznávání osob pomocí dat z kamery a mmWave radaru pro spolehlivější detekci a zobrazení informací o objektu (vzdálenost) na základě dat z milimetrového radaru. Následuje přesnější zadání tohoto úkolu, pro které jsem se rozhodl na základě analýzy současných řešení.

3.4 Technické zadání

Technické zadání, pro které jsem se rozhodl, popisuje požadované vlastnosti řešení úkolu detekce osob za pomoci dat senzorů. Znění technického zadání je pak následující:

- Využijte konvoluční neuronovou síť (CNN) k fúzi informací z milimetrového radaru a monokulární kamery.
- CNN bude dále použita k detekci osob ve vstupních datech.
- Pro daný úkol detekce osob navrhnete pro fúzi vhodné rozmístění senzorů.
- Za účelem sběru dat pak využijte sériové rozhraní milimetrového radaru a USB rozhraní běžně dostupné webové kamery.
- Nad binárními daty z milimetrového radaru provedte syntaktickou analýzu.
- Zvolte a implementujte vhodnou transformační metodu pro reprezentaci dat z radaru v obrazových datech kamery.
- Navrhnete architekturu konvoluční neuronové sítě a za použití vhodného modulu ji realizujete.
- Vytvořte dataset vhodný pro daný případ ověření – pozitivní a negativní vzorky dat.
- Provedte trénování CNN, využijte při tom metody, které pomáhají zabránit přeučení.

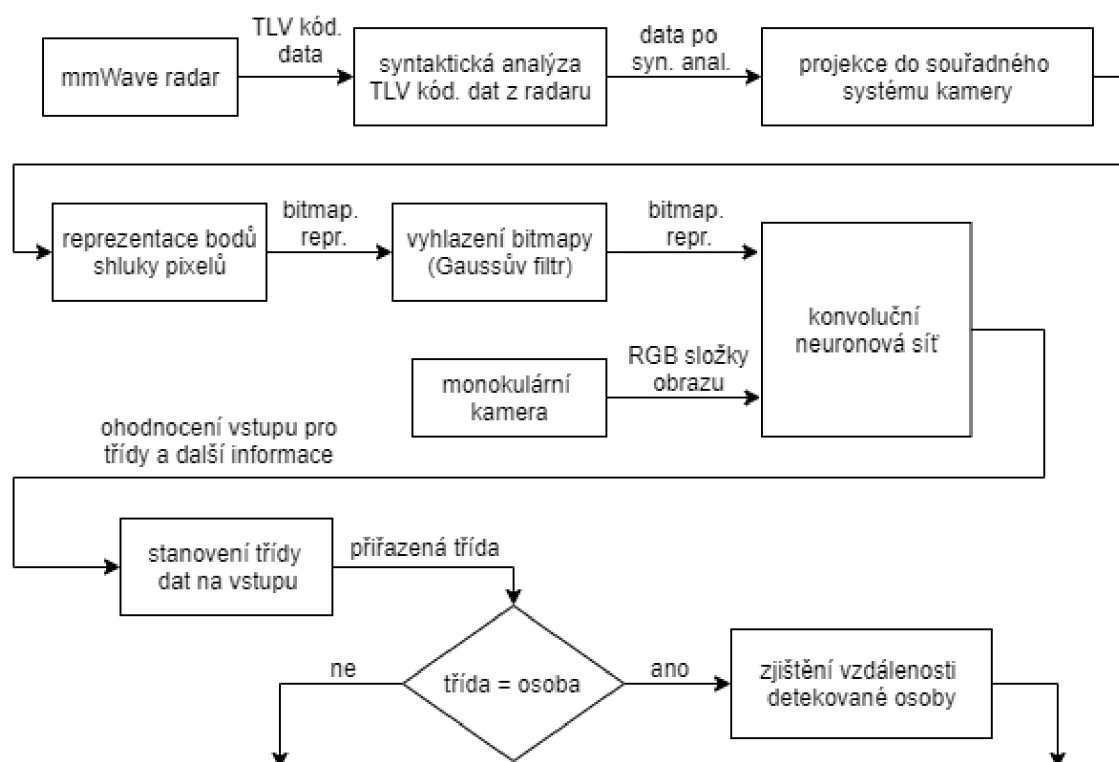
- Pro detekovanou osobu určete vzdálenost od senzorů pomocí vhodného způsobu vyhledání osobě odpovídajících dat z milimetrového radaru.
- Výstupem systému bude obraz z kamery s rámečkem označujícím detekovanou osobu a informací o její vzdálenosti.

Následující kapitola se věnuje popisu řešení systému využívajícího fúzi informací milimetrového radaru a kamery za účelem detekce osob a zobrazení její vzdálenosti od mmWave radaru.

Kapitola 4

Popis vlastní práce

Tato kapitola se věnuje popisu řešení navrženého systému. Popisuje použité metody a postupy, věnuje se sběru dat, způsobu ověření a analýze výsledků, které byly získány. Snaží se při tom vycházet z již zjištěných poznatků, které využívá. Na následujícím obrázku se nachází blokové schéma reprezentující proces zpracování dat z milimetrového radaru a kamery 4.1, jednotlivé bloky vyjadřující části řešení jsou do jisté míry logickým rozvržením této kapitoly. Pro implementaci systému byl vybrán objektově orientovaný jazyk Python kvůli rozmanitosti použitelných modulů a flexibilitě.



Obrázek 4.1: Blokové schéma systému

4.1 Sběr dat z milimetrového radaru

Způsob fungování 2.3 a možnosti snímání 2.4 milimetrového radaru jsou probrány v kapitole 2. Čtenáři by tedy již měl být známý způsob fungování milimetrového radaru (mmWave radaru) a možnosti, které jsou k dispozici pro sběr dat. Je tedy třeba zvolit vhodný demonstrační příklad pro zvolený případ užití (detekce osob) a získat poskytovaná data.

Z řady dostupných demonstračních příkladů, které výrobce poskytuje, se jako nejlepší jeví některé pro počítání osob v místnosti. Tyto demonstrační příklady poskytují kromě mračna bodů i další užitečné informace. V rámci radaru totiž běží i algoritmy pro detekci a sledování pohybu osob. V případě, že radar rozezná osobu, přidává k mračnu bodů i informace jako:

- umístění osoby pomocí kartézských souřadnic (x, y) ,
- rychlosti pohybu osoby ve směru osy x a y ,
- akceleraci ve vztahu k osám x a y ,
- id osoby a přiřazení detekovaných bodů z mračna bodů k osobám.

Jelikož byl pro tuto práci využit typ mmWave radaru IWR6843ISK od firmy Texas Instruments (TI), použil jsem odpovídající demonstrační příklad pro počítání osob „68xx ISK – People Counting“ obsaženou v toolkitu verze 4.1.0.

Data z radaru jsou přenášena pomocí asynchronní komunikace typu UART přes rozhraní USB. Musí se tedy nad těmito daty provést syntaktická analýza. Je třeba vycházet z dokumentace výše zmíněného demonstračního příkladu [3] a na základě kódování data reprezentovat. Demonstrační příklad používá kódování Type-length-value (TLV), které pracuje s bloky dat. První součástí, radarem poslaného paketu, je hlavička obsahující základní informace o přenosu, synchronizační řetězec, paketu, dále pak checksum a taky počet TLV bloků obsažených v daném paketu. Tyto bloky se pak dělí na několik typů, podle toho jestli blok přenáší data o mračnu bodů, detekované osobě, nebo o přiřazení bodů k odpovídajícímu shluku (osobě). Popsaný paket vypadá jako na obrázku 4.2.



Obrázek 4.2: Rozdělení paketu radaru [3]

Aby však k přenosu vůbec došlo, je třeba zahájit komunikaci mezi počítačem a radarem. Je třeba nahrát konfiguraci radaru a poté přijímat binární data, která pak musí být dále zpracována. K tomuto účelu mé řešení využívá upravenou část skriptu dostupného v rámci fakulty informačních technologií VUT v Brně, jenž poskytuje data pro následnou syntaktickou analýzu.

Výsledkem syntaktické analýzy paketu odeslaného mmWave radarem jsou data právě jednoho měření. Syntaktická analýza je prováděna za pomoci modulu `Struct`, která dokáže pracovat s binárními daty a interpretovat je jako určitý datový typ. V první řadě je však třeba najít začátek přenášeného paketu. Toto se dá realizovat pomocí výše zmíněného synchronizačního řetězce, což je v kontextu přenosu unikátní posloupnost znaků značící začátek daného paketu. Jakmile je k dispozici pozice synchronizačního řetězce (začátku paketu), je

možné podle dokumentace [3] interpretovat jednotlivé byty. Pokud mají být první čtyři byty po synchronizačním řetězci reprezentovány jako datový typ „uint32“, můžeme tato data pomocí výše zmíněného modulu interpretovat a zjistit v tomto případě verzi demonstračního příkladu [3]. Stejným způsobem lze získat detekované body a objekty z jednotlivých bloků TLV.

4.2 Popis hardware a umístění senzorů v prostoru

Tato část je umístěna po části **Sběr dat z milimetrového radaru** kvůli návaznosti na zvolený demonstrační příklad použitý pro sběr dat z milimetrového radaru. Je totiž třeba radar nastavit do určité výšky, pod daným úhlem atd. Pro realizaci snímání byl použit milimetrový radar IWR6843ISK od firmy Texas Instruments (TI) a webová kamera Logitech Pro Stream Webcam C922 PRO.

Demonstrační příklad pro sledování počtu osob v místnosti [3] doporučuje umístit radar do výšky 1,5 m až 2,5 m a kvůli kalibraci kamery a radaru je vhodné kameru umístit do podobné výšky. Kamera by totiž na scénu měla nahlížet shora. Bylo tedy třeba umístit oba senzory na stativ a zkonstruovat prostředek umožňující situovat kameru a radar vedle sebe na pohyblivou hlavu stativu. Kvůli maximální výšce podstavce nebylo možné umístit stativ na zem, bylo třeba najít alternativu. Tou je umístění stativu na stůl o standardní výšce 75 cm a tím získání efektivní výšky asi 215 cm, což je vyhovující. Je tedy stále možné jednoduše umístit tuto konfiguraci na jiné místo pro sběr dat a zachovat perspektivu kvůli kalibraci senzorů (více viz sekce 4.3).



Obrázek 4.3: Umístění senzorů na stativu – verze 1

Během testování však byla u výše zmíněného rozmístění nalezena vada. Radar totiž produkuje mračno bodů v rovině, která je téměř totožná s rovinou danou dvěma osami soustavy souřadnic kamery. Je tedy možné zobrazit body z mmWave radaru do obrazu kamery pouze jako úsečku. Bylo tedy na základě tohoto zjištění třeba rozmístění senzorů změnit vyobrazeným způsobem 4.4.



Obrázek 4.4: Umístění senzorů na stativu – verze 2

Milimetrový radar se v pozměněné konfiguraci 4.4 nachází o 90 *cm* níž než kamera a svírá vzhledem k podložce kolmý úhel. Toto rozvržení již umožňuje zobrazení bodů radaru do obrazu kamery jako 2D objektu. Kamera totiž na rovinu danou mračnem bodů shlíží a dokáže tak projekcí zobrazit body, které jsou v soustavě souřadnic radaru v zákrytu. Zároveň si toto rozmístění zachovává výhody předešlé konfigurace. Oba senzory se totiž stále nacházejí na stejném stativu, a tak je možné řešení přemístit bez nutnosti kalibrace. Parametry rozmístění senzorů pro druhou verzi, které nebyly zmíněny v tomto odstavci, zůstaly zachovány.

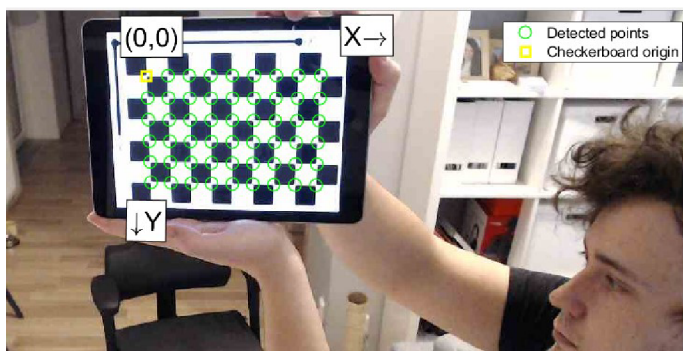
Nevýhodou druhé verze rozvržení je nastavení demonstračního příkladu pro počítání osob. To totiž s umístěním radaru do kolmé pozice vůči podložce nepočítá. Radar sice dokáže detekovat i v této pozici objekty v zákrytu, dochází zde však k produkci bodů odpovídajících neexistujícím objektům. Tento fakt bude třeba zohlednit při měření vzdálenosti osoby od senzorů, více viz sekce 4.7.

4.3 Transformace dat milimetrového radaru

Tato část pojednává o transformaci bodů z milimetrového radaru do soustavy souřadnic kamery. Je tedy třeba získat transformaci, která dokáže na základě vstupních dat nebo informací o prostoru transformovat body z radaru do kamery. Lze ji ilustrovat obrázkem 2.11. Tento problém řeší již existující práce [23, 20, 9] a podrobněji je probrán v části 2.8. Následující část se tedy věnuje spíše praktickému postupu kalibrace senzorů, než vysvětlování metody samotné. Ta je totiž podobná postupu použitému ve studii [9].

Nejprve byla použita kalibrační metoda využívaná v pracích [23, 20], která transformační matici pro převod bodů z radaru do soustavy souřadnic kamery počítá pomocí metody nejmenších čtverců na základě souřadnic odpovídajících bodů (kamera, radar). Tento přístup se však ukázal jako nevyhovující, jelikož nedocházelo ke správné projekci vzdálených bodů radaru do obrazové roviny kamery, a proto byla použita metoda popsaná v práci [9], kde je využita pro rekonstrukci 3D scény na základě dat z mmWave radaru a kamery, více v části 2.8.

Pro stanovení dané transformace, je třeba získat vnější (umístění senzorů v prostoru) a vnitřní (charakteristiky kamery) parametry. Stanovit vnější parametry je jednoduché, ty vycházejí z rozvržení senzorů v prostoru, a tedy z informací z části 4.2. Naopak vnitřní parametry nelze jednoduchým způsobem pozorovat. Na štěstí k tomuto účelu existuje aplikace od společnosti MathWorks pro kalibraci kamery, která na základě fotografií šachovnice v prostoru detekuje ohniskové vzdálenosti (f_x a f_y) a hlavní středový bod (p_x a p_y) a tedy potřebné vnitřní parametry. Příkladem fotografie použité pro stanovení vnitřních parametrů kamery je obrázek 4.5.



Obrázek 4.5: Fotografie šachovnice ta účelem kalibrace kamery

Po získání všech parametrů je možné provést kalibraci. Je třeba nalézt matici rotace R a vektor translace t kamery v prostoru pro definici pozice kamery v prostoru. K tomuto účelu lze využít úhel, který osa y kamery (osa výšky umístění bodu) svírá s odpovídající souřadnou osou světové soustavy souřadnic. Tuto hodnotu je pak možné použít za pomoci patřičného vztahu k nalezení matice rotace na základě úhlu. Umístění kamery v prostoru je definováno v části 4.2. V případě rozmístění senzorů nad sebou bude třeba translace ve směru osy y (nahoru, nebo dolů).

V této fázi je již možné sestavit matici P_{cw} , která definuje vztah soustavy souřadnic kamery a světové soustavy souřadnic. Jelikož se však jedná o vyjádření pozice kamery v prostoru, je třeba matici invertovat a tím získat matici P_{wc} vyjadřující pozici počátku soustavy souřadnic světa v systému souřadnic kamery. Před inverzí je třeba matici P_{cw}

složenou z matice rotace R (o velikosti 3×3) a vektoru translace kamery t doplnit o homogenní souřadnice. Popisovaný vztah mezi uvedenými maticemi je uveden níže 4.1.

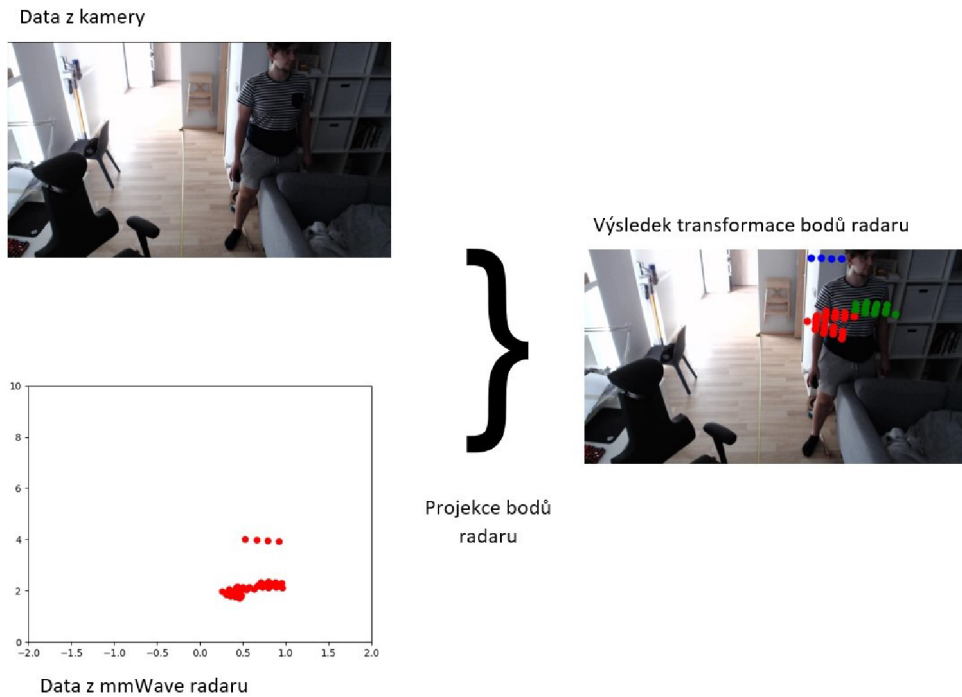
$$P_{cw}^{-1} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}^{-1} = P_{wc} \quad (4.1)$$

Dalším krokem je pak projekce bodů ze soustavy souřadnic kamery do systému souřadnic obrazové roviny kamery. Toto je možné provést díky zjištěným vnitřním parametrům kamery. Jedná se o středovou projekci, která lze vyjádřit pomocí vztahů 4.2 a 4.3.

$$u = f_x \times \frac{x_w}{z_w} + p_x \quad (4.2)$$

$$v = f_y \times \frac{y_w}{z_w} + p_y \quad (4.3)$$

kde souřadnice u a v jsou souřadnice bodu na obrazové rovině kamery. V praxi je pak tedy možné na základě zadaných údajů jednoduše zkalibrovat senzory při změně jejich rozmístění v prostoru. Příkladem průběhu projekce bodů na obrazovou plochu může být obrázek 4.6. Body jsou barevně rozlišeny, jelikož jsou rozděleny na shluky, které jsou použity pro zjišťování vzdálenosti detekovaných osob, více viz 4.7.



Obrázek 4.6: Aplikace transformace dat z mmWave radaru a zobrazení do obrazu kamery

4.4 Zpracování dat milimetrového radaru pro klasifikaci

Tato část se zabývá preprocessingem dat mmWave radaru pro použití při klasifikaci pomocí konvoluční neuronové sítě. V tuto chvíli jsou data z milimetrového radaru transformována do soustavy souřadnic obrazové roviny kamery, dalo by se však tvrdit, že jsou to pouze jednotlivé pixely, které dobře nerepresentují informace z mračna bodů radaru. Rozlišitelnost objektů pomocí milimetrového radaru je totiž nižší než rozlišení kamery.

V dokumentaci použitého demonstračního příkladu [3] lze nalézt rozlišitelnost objektů vedle sebe, která činí pro demonstrační příklad pro počítání osob asi 1,4 stupně. Použité rozlišení kamery je 1280 pixelů. Úhel zorného pole kamery je 78 stupňů a milimetrového radaru 70 stupňů. To znamená, že maximální počet bodů vedle sebe je asi 1148 v případě kamery a 50 bodů pro radar a bod radaru by tedy měl být v horizontální ose reprezentován asi 18 pixely, aby byl zachován význam dat z mmWave radaru. Je také vhodné do reprezentace bodů radaru zahrnout další informace poskytované radarem jako je SNR hodnota bodů.

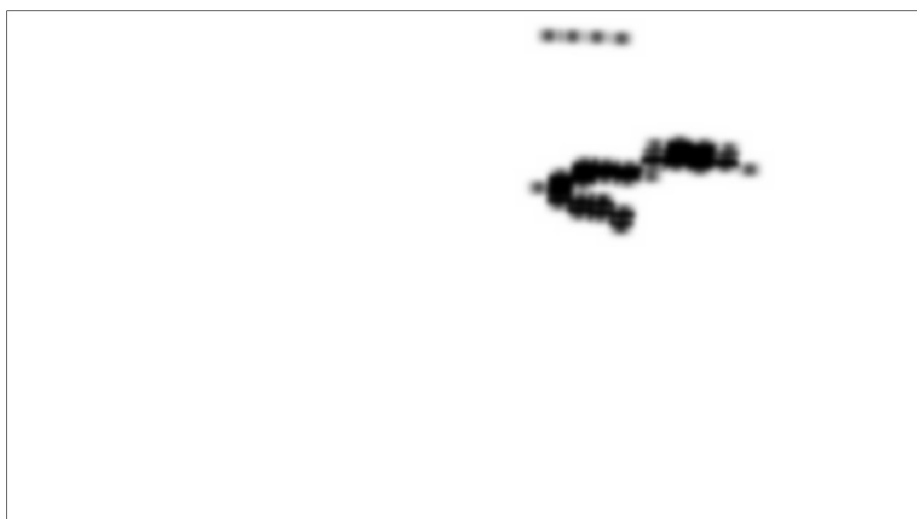
Je tedy třeba najít způsob, jak body co nejlépe reprezentovat pomocí obrazových dat. Je třeba informaci o bodech z mmWave radaru rozšířit do okolí. Tato práce problém šíření řeší pomocí aplikace 2D Gaussovského rozložení na každý z bodů radaru. Toto normální rozložení je vynásobeno hodnotou SNR, a tím se tato informace o bodech neztrácí. Při tom jsou hodnoty SNR namapovány na hodnoty, kterých může nabývat černobílý obrázek a dá se tedy jednoduše zobrazit. Nejvyšší hodnota SNR mračna bodů je namapovaná na nejvyšší možnou hodnotu pixelu.

Velikost filtru vychází z části z úvahy výše a z části z experimentálního stanovení parametrů. Jako vyhovující parametry, pro 1D rozdělení definující 2D rozdělení, se ukázaly parametry dva. Jeden pro rozdělení ve směru horizontálním a druhý pro rozdělení ve směru vertikálním. S tím, že rozdělení ve směru vertikálním, má menší velikost, než ve směru horizontálním. Je to kvůli provedené transformaci bodů radaru. Ty jsou totiž promítány ve směru vertikálním s menšími rozestupy a při použití stejných hodnot pro obě 1D rozložení docházelo ke špatným výsledkům. Po aplikaci jednotlivých jader (filtrů) na body z radaru, je ještě celá bitmapa zpracována pomocí vyhlazovacího algoritmu (Gaussův filtr).

Body vyobrazené v obrázku 4.7 jsou pak reprezentovány jako je tomu na obrázku 4.8. Na obrázku 4.8 je také možné pozorovat vliv hodnoty SNR na reprezentaci bodů. Některé jsou totiž reprezentovány šedou barvou (nízká hodnota SNR) a některé černou (vysoká hodnota SNR).



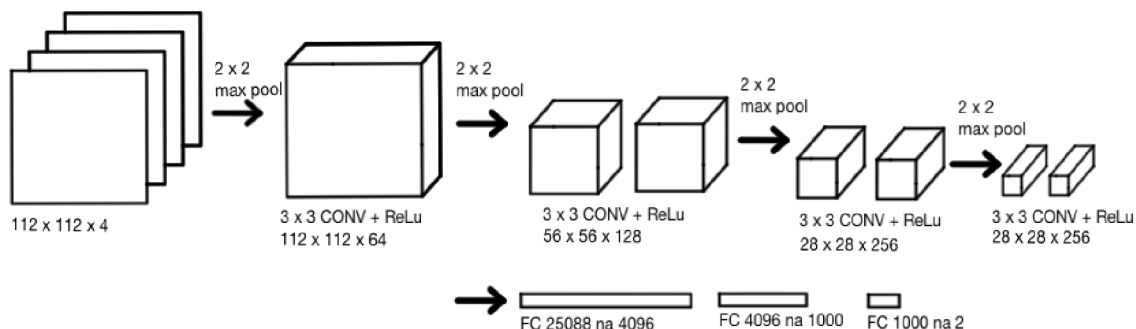
Obrázek 4.7: Mračno bodů radaru zobrazeno v obraze



Obrázek 4.8: Repräsentace mračna bodů radaru pomocí bitmapy

4.5 Návrh konvoluční neuronové sítě

Návrh konvoluční sítě vyplývá z architektury VGGNet použité pro ukázkou v sekci 2.7. Vyhodnocuje jednotlivé části dat za účelem detekce osob a předchází jí segmentace pomocí metody okénka. Oproti architektuře VGGNet konvoluční neuronová síť pracuje se čtyřmi vrstvami dat místo tří, přidává se k ní totiž vrstva vytvořená za pomoci milimetrového radaru. Navržená konvoluční neuronová síť je implementována pomocí Python modulu – PyTorch a její schéma se nachází na obrázku 4.9.



Obrázek 4.9: Architektura použité konvoluční neuronové sítě

Architektura navržené sítě vychází z dat o rozměrech 112×112 pixelů místo vstupní velikosti obrázků VGGNet (224×224 pixelů). Skládá se z konvolučních vrstev o velikosti 3×3 , s padding 1 a stride 1. Dále pak jsou použity aktivační vrstvy typu ReLu a max pooling vrstvy o velikosti 2×2 . Jako klasifikační vrstvy jsou použity tři plně propojené vrstvy, jejichž výsledkem jsou dvě třídy značící, zda jsou vstupní data osoba, nebo nikoliv.

Po provedení klasifikace za pomoci plně propojených vrstev je třeba provést rozhodnutí, do které ze zkoumaných tříd obrázků náleží. Je možné pouze zvolit třídu s nejvyšším skóre, to by však s velkou pravděpodobností vedlo k velkému množství chyb typu II (false positives). Je tedy vhodné nastavit pro rozpoznávané třídy určitý rozhodovací práh, který musí být překročen, aby byla vstupní data zařazena do dané třídy.

Jelikož byla CNN trénována na malém datasetu, bylo třeba zabránit přeučení, více viz 4.6. K tomuto účelu je možné využít různé metody. V navržené architektuře to jsou dropout [16] vrstvy u plně propojených vrstev a batch normalization [21] vrstvy u konvolučních vrstev.

4.6 Trénování konvoluční neuronové sítě

Tato část popisuje proces trénování konvoluční neuronové sítě. Pro tuto fázi bylo nutno shromáždit data, neexistuje totiž dataset, který by zahrnoval jak data z kamery, tak z milimetrového radaru. Sběr dat tedy probíhal za pomoci dobrovolníků z řad rodiny a přátel.

Dataset byl rozšířen pomocí různých transformací vstupních dat. Byly použity následující metody: otáčení o náhodný úhel, ořezávání a horizontální převrácení. U neuronových sítí, zvláště pak u těch, které jsou trénovány na malém datasetu (jako je tato), může docházet k přeučení. Přeučená neuronová síť správně funguje pouze v případech, kdy jsou vstupní data velmi podobná trénovacím, schází jí schopnost generalizace. Možností jak přeučení zabránit je mnoho, v případě této práce byly využity metody *Batch Normalization* [21] a *Dropout* [16]. První zmíněná metoda zlepšuje učící schopnost neuronové sítě a snižuje důraz kladený na inicializaci vah. Podle studie [21], kde byla navržena, také pomáhá s generalizací. Dropout je metoda, která při každém průchodu s určitou pravděpodobností zruší spojení mezi některými neurony sítě, a tím zajistí, že jednotlivé vrstvy na sobě nejsou příliš závislé a pomáhá tak s generalizací.

Bylo také třeba jednotlivé trénovací vzorky izolovat od okolí. K tomuto účelu byl vytvořen skript, který uživateli dává možnost z nashromážděných dat vybrat jednotlivé sekce, které jsou pak určeny právě k tréninku CNN. Bylo třeba nashromáždít vzorky jak osob, tak prostředí – pozitivní a negativní.

Samotné trénování je uskutečněno za pomoci ztrátové funkce, která hodnotí výsledky na základě křížové entropie. Hodnocení probíhá za pomoci vztahu 4.4, kde $p(x)$ jsou správné třídy pro jednotlivé vzorky a $q(x)$ jsou předpovězené třídy pomocí CNN. Výsledkem křížové entropie dokonalé neuronové sítě, která by předpověděla všechny rozpoznávané vzorky správně, by byla nula, čím se od sebe distribuce naopak více liší, tím ztrátová funkce nabývá větší hodnoty.

$$H(p, q) = - \sum_{\forall x} p(x) \log(q(x)) \quad (4.4)$$

K aplikaci vypočtené hodnoty ztrátové funkce je pak využit optimalizátor implementující stochastické klesání podle gradientu. Jediný rozdíl mezi algoritmem klesání podle gradientu (GD) (více informací v části 2.7) a stochastického klesání podle gradientu (SGD) je pak způsob výběru vzorků v jednotlivých iteracích. U GD jsou pro každou úpravu vah CNN použity všechny trénovací vzorky, u SGD je to jen část (dávka). V mém případě nejlépe funguje dávka o velikosti 10, ta je z hlediska využití paměti dostatečně malá a z hlediska generalizace naopak dostatečně velká. Menší velikost dávky také není vhodné použít kvůli použité metodě *Batch Normalization* [21], která vyžaduje pro správné fungování minimálně uvedenou velikost 10 vzorků. Vzorky trénovacích dávek jsou s každou iterací náhodně měněny.

Jelikož je dataset, který pro trénink používám, malý, obávám se, že neuronová síť nebude schopna dobře generalizovat. Tato domněnka je prověřena v části **Testování systému**.

4.7 Stanovení vzdálenosti detekované osoby

Poté, co je obraz konvoluční neuronovou sítí zpracován, je možné o objektech zjistit další informace za pomoci dat milimetrového radaru, jako třeba vzdálenost detekované osoby. Tento údaj je pak možné přiřadit detekovanému objektu, v případě této práce např. osobě.

Je tedy nutné v oblasti, kde byla osoba detekována vyhledat příslušný shluk z mračka bodů milimetrového radaru a díky němu zjistit vzdálenost detekované osoby. Tento krok je zapotřebí, protože se v daném regionu mohou vyskytovat body, které osobě nepatří. Může to být odraz od reflexivního povrchu, jiná osoba, zvíře apod. Z toho důvodu mé řešení využívá shlukovací algoritmus za účelem rozdělení bodů milimetrového radaru do skupin, které s vysokou pravděpodobností patří stejnému objektu. Příkladem rozdělení mračka bodů na jednotlivé shluky je obrázek 4.10, kde jsou shluky rozlišeny barvou bodů. Po vyhodnocení vstupních dat a nálezů oblasti, ve které byla detekována osoba dochází k výběru shluku, který je v daném segmentu zastoupen nejvíce, a dále použit pro výpočet vzdálenosti detekované osoby.



Obrázek 4.10: Vstupní obrázek z kamery obohacený o body radaru rozdělených do shluků

Vybraný shluk obsahuje kromě souřadnic v obrazové rovině i souřadnice v světové soustavě souřadnic, ze kterých je možné vyjádřit vzdálenost nalezeného objektu od mmWave radaru. Zpracovávaný shluk bodů však může obsahovat body, které k osobě nepatří, a proto je vzdálenost objektu stanovena pomocí N k senzoru nejbližších bodů vybraného shluku. V části testování se toto řešení ukázalo jako nejpřesnější více viz sekce 4.8.

Pro zjištění vzdálenosti by bylo možné využít informace v rámci konvoluční neuronové sítě. Nebylo by pak třeba odpovídající část mračka bodů milimetrového radaru vyhledávat výše popsanou metodou. Místo toho by byly potřebné body milimetrového radaru známy a výpočet vzdálenosti by tak byl triviální.

Výsledkem detekce a následného stanovení vzdálenosti osoby je pak obrázek obohacený o rámeček značící region, kde se vyskytuje osoba, kruh značící, kde byla na osobě měřena vzdálenost a vzdálenost tohoto člověka v centimetrech. Příkladem může být obrázek 4.11.



Obrázek 4.11: Vstupní obrázek z kamery obohacený o rámeček značící přítomnost osoby a její vzdálenost

4.8 Testování systému

Účelem provedeného testování bylo prověření jednotlivých vlastností realizovaného systému. Těchto parametrů je však mnoho, proto je třeba ověřit ty nejdůležitější:

- zda systém detekuje osoby,
- jestli systém pracuje ve zhoršených světelných podmínkách,
- přesnost měření vzdálenosti detekované osoby.

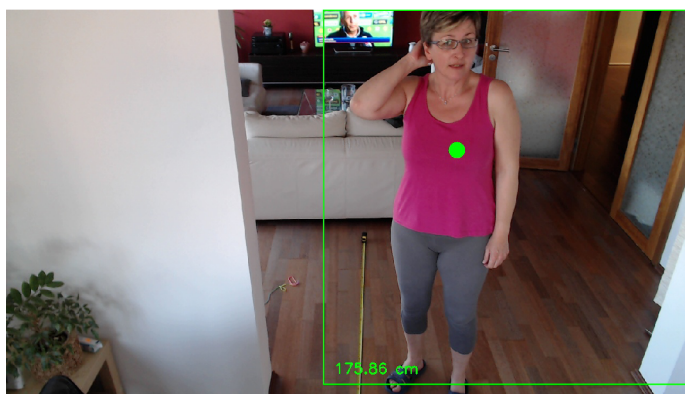
Pro pochopení výsledků testů je třeba znát výstup systému. Tím jsou jednotlivé obrázky níže, zelený rámeček vždy označuje část obrázku, ve které se nachází osoba a v levém horním rohu je umístěna vzdálenost detekovaného člověka. Zelený kruh pak vyobrazuje místo, kde byla na osobě měřena vzdálenost.

Pro testování detekce osoby byla použita vstupní data, která byla pro systém neznámá. To znamená data pořízená v neznámém prostředí s neznámými osobami. Příkladem výsledku zpracování těchto dat jsou obrázky 4.12 a 4.13. Další výsledky testování jsou obsaženy v následující příloze této práce A.1.

Aby bylo možné výsledek této části testování ohodnotit, je třeba stanovit kritéria, která budou dále použita pro vyjádření úspěšnosti. Těmi jsou:

- detekce osoby
- a správnost umístění detekčního rámečku.

Detekcí osoby se pak rozumí umístění rámečku do oblasti obrázku, kde se skutečně nachází osoba. Správnost umístění rámečku je dána podílem plochy osoby nacházející se v detekčním rámečku. Pokud se nachází v rámečku většina plochy osoby, je umístění rámečku správné. Je tedy možné aby byla detekce úspěšná, i když se uvnitř rámečkem ohraničené oblasti nenachází některá z končetin, část hlavy apod. Zároveň je však pro to aby bylo umístění rámečku považováno za úspěšné třeba, aby rámeček nepokrýval příliš velký region, a byla to tedy pouze oblast, která je potřeba pro zahrnutí celé detekované osoby.

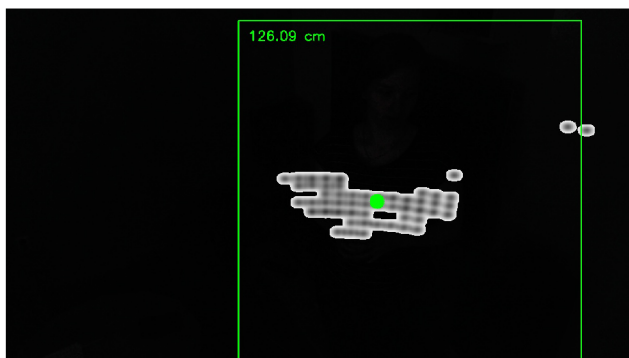


Obrázek 4.12: Příklad výstupu systému

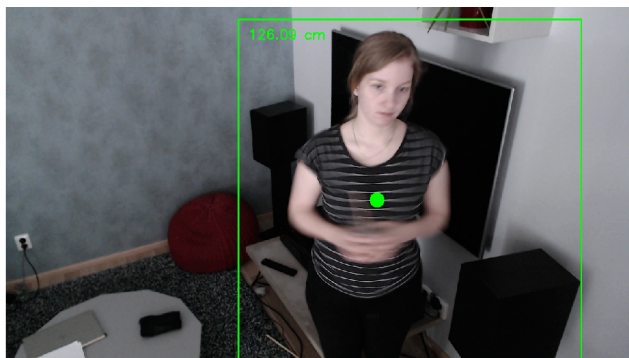


Obrázek 4.13: Příklad výstupu systému

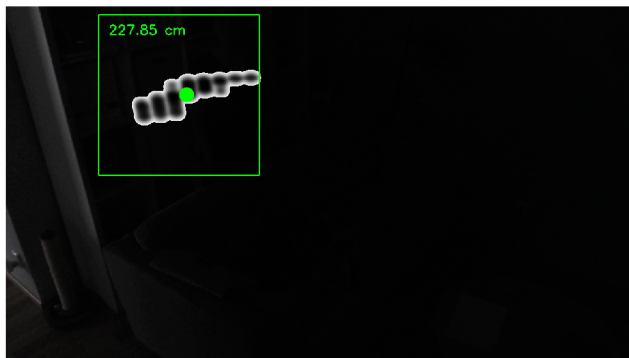
Další testovanou vlastností, která byla třeba ověřit, bylo fungování ve zhoršených světelných podmínkách. V těchto situacích by měla CNN spoléhat hlavně na data z milimetrového radaru a podpořit tak rozpoznávání v situaci, kdy kamera zaostává. Kvůli tomu jsou na následujících obrázcích vyobrazeny i body z milimetrového radaru, obrazová data z kamery totiž nemají velkou vypovídací hodnotu. Jedná se o obrázky 4.14 a 4.16. Pro srovnání jsou zde poskytnuty i odpovídající obrázky v dobrých světelných podmínkách 4.15 a 4.17. Zde je kritérium úspěchu pouze, zda byla osoba detekována ve správné části vstupních dat.



Obrázek 4.14: Výsledek pro testovací vstup 1 ve špatných světelných podmínkách



Obrázek 4.15: Výsledek pro odpovídající test. vstup 1 v dobrých světelných podmínkách



Obrázek 4.16: Výsledek pro testovací vstup 2 ve špatných světelných podmínkách



Obrázek 4.17: Výsledek pro odpovídající test. vstup 2 v dobrých světelných podmínkách

Posledním bodem testování byla přesnost detekce vzdálenosti objektu od senzorů. Ta je zjišťována způsobem popsaným v části 4.7 a určuje vzdálenost detekované osoby. Pokud se v oblasti určené rámečkem vyskytuje více osob, je zjištěna vzdálenost člověka, ke kterému náleží nejvíce bodů z mráčna bodů radaru. Je to kvůli umístění milimetrového radaru, kdy dochází ve vyšší míře k detekci bodů, které nutně nemusí patřit k žádné osobě. Není tedy vhodné v případě nálezu většího počtu shluků v detekované oblasti uvádět vzdálenosti pro každý z nich. Je totiž možné, že některé shluky nepatří k žádné osobě a neměla by tedy u nich být určena vzdálenost, více viz 4.7.

Způsobem testování přesnosti měření vzdálenosti pak bylo porovnání systémem stanovených vzdáleností s referenčním měřením vzdálenosti osoby od senzorů měřidlem. Jelikož nebylo možné referenční hodnoty získat s přesností, se kterou jsou poskytovány systémem, jsou výsledky uvedeny s přesností na centimetry.

Výsledky testování

Výsledky testování jsou uvedeny v tabulce 4.18. Pro každou z ověřovaných vlastností systému bylo použito 20 testovacích vstupů. Některé z použitých testovacích vstupů a jim náležících výsledků jsou čtenáři k nahlédnutí dostupné v příloze A této práce.

Testování detekce osob	Hodnoty
Úspěšnost detekce osob	100%
Úspěšnost umístění detekčního rámečku	75%
Testování detekce osob ve zhoršených podmínkách viditelnosti	Hodnoty
Úspěšnost detekce osob ve zhoršených světelných podmínkách	95%
Testování přesnosti měření vzdálenosti osoby od senzoru	Hodnoty
Maximální záporná chyba měření vzdálenosti systému	-19 cm
Maximální kladná chyba měření vzdálenosti systému	10 cm
Směrodatná odchylka měření	9 cm

Obrázek 4.18: Výsledky testování

4.9 Zhodnocení výsledného řešení systému pro detekci osob

Navržený a dále realizovaný systém je schopen detekce osob a stanovení jejich vzdálenosti v prostoru. Jelikož však pro fúzi a klasifikaci využívá konvoluční neuronovou síť, trpí nedostatkem trénovacích dat. Pokud by byl použitý dataset rozšířen, spolehlivost a schopnost generalizace tohoto řešení by vzrostla. V případě, že by dále byla trénovací data obohacena o jiné typy objektů (cyklisty, automobily atd.), bylo by možné podobný systém využít k jejich rozpoznávání, a tím např. pomáhat v různých dopravních situacích, kde kamera v nepříznivých podmínkách přestává fungovat.

Kromě nedostatku trénovacích dat je možné u navrženého řešení pozorovat další problém a tím je rychlost detekce. Ta je totiž omezena použitou metodou hledání oblastí zájmu – metodou okénka. Pokud by tedy měl být realizovaný systém vylepšen, bylo by vhodné místo této metody použít např. algoritmus selective search, nebo se inspirovat jednorůchodovými klasifikačními metodami a tím celý proces detekce přenechat konvoluční neuronové síti.

V této práci nebyla zmíněna možnost využití 3D milimetrového radaru, který dokáže detekovat i výšku bodů v prostoru a je ho tedy možné využít podobně jako 2D milimetrový radar, s tím rozdílem, že jej lze umístit na stejné místo jako kameru. Výsledkem transformace mračna bodů radaru do soustavy souřadnic kamery, v případě umístění na identické místo, již totiž není úsečka, jako tomu bylo při nesprávném rozmístění senzorů, které je popsáno v části 4.2, je to 2D množina bodů. Ta pak lze jednoduše zkalibrovat s kamerou a ještě snadnějším úkolem je projekce těchto bodů do obrazové roviny kamery.

Proč tedy tří-dimenzionální milimetrový radar nepoužít?

Ze dvou důvodů: 3D milimetrové radary jsou dražší a pokud je pro určitý případ užití 2D milimetrového radaru dostačující, není třeba 3D variantu radaru použít. Pokud je tedy třeba produkovat výrobek ve velkých sériích a je 2D mmWave radar dostačující, je možné při použití 2D radaru ušetřit. Existují také další senzory, které pracují pouze s 2D daty. Podobný způsob fúze, jako je realizován v této práci, lze tedy použít pro více typů senzorů, např. pro fúzi kamery a 2D lidarů.

Výše zmíněné však neznamená, že by 3D milimetrový radar neměl být pro účely fúze s kamerou a následné rozpoznávání využíván. Naopak, v situacích, kdy není možné umístit senzory na různá místa, nebo není možné v daném prostoru použít kolmo k podložce umístěný radar, je nutné využít 3D radar. Cena této varianty milimetrového radaru také klesá a měla by tedy být dále více využívána. Vhodným pokračováním této práce by bylo porovnání systémů využívající 2D a 3D radar. Co se týče rozpoznávání objektů, 3D milimetrový radar má oproti 2D variantě výhodu v tom, že je díky produkováním tří-dimenzionálním datům možné body do prostoru kamery umístit s mnohem vyšší přesností, ne pouze jakýsi průřez objekty. Tento fakt by mohl pomoci schopnosti systému rozpoznávat různé objekty ve špatných podmínkách viditelnosti.

Kapitola 5

Závěr

Cílem této práce bylo prozkoumat vlastnosti systému využívajícího fúze informací z milimetrového (mmWave) radaru a kamery k rozpoznávání objektů na vhodném příkladu za použití konvoluční neuronové sítě.

Tento záměr byl splněn a systém dokáže detekovat osoby za pomoci dat z obou senzorů a přiřadit jim vzdálenost zjištěnou na základě dat z mračna bodů mmWave radaru. Vybraným příkladem pro ověření tedy byla detekce osob v prostoru a následné získání informace o vzdálenosti člověka od senzorů.

Pro návrh realizovaného systému bylo třeba prostudovat dostupnou literaturu týkající se především fúze sensorických dat a konvolučních neuronových sítí. Pro demonstraci fúze kamery a mmWave radaru byl zvolen vhodný příklad k ověření vlastností a navržen možný postup za pomoci konvoluční neuronové sítě. Tento návrh bylo třeba vhodným způsobem implementovat, byly proto prozkoumány dostupné moduly a knihovny, které mohou být využity pro realizaci jednotlivých částí systému. Funkčnost navrženého systému pak byla ověřena na testovacích datech, které prověřily jednotlivé aspekty řešení.

Práce mě naučila mnohé. Bylo třeba provést syntaktickou analýzu dat z mmWave radaru, což je nízkoúrovňový úkol. Na druhou stranu jsem musel navrhnout, implementovat a trénovat konvoluční neuronovou síť. V této práci se tedy bylo třeba vypořádat s nejrůznějšími problémy, které jsem musel pochopit, prostudovat a vyřešit.

Možným pokračováním této práce je využití tří-dimenzionální varianty mmWave radaru pro srovnání fúze navržené v této práci a fúze využívající právě 3D radar. Vhodným pokračováním by pak také bylo rozšíření datasetu pro trénování konvoluční neuronové sítě a s tím související možnosti detekce dalších objektů. Zajímavým případem by bylo třeba sledování dopravní situace – detekce automobilů, chodců, cyklistů apod.

Literatura

- [1] *Signal-to-Noise Ratio (SNR) and Wireless Signal Strength* [online]. San José, Spojené státy americké: Cisco Systems, 2018 [cit. 2020-05-15]. Dostupné z: [https://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/Signal-to-Noise_Ratio_\(SNR\)_and_Wireless_Signal_Strength](https://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/Signal-to-Noise_Ratio_(SNR)_and_Wireless_Signal_Strength).
- [2] *Comparing Nearest Neighbors with and without Neighborhood Components Analysis* [online]. Scipy (Scientific computing tools for Python), 2019 [cit. 2020-04-25]. Dostupné z: https://scikit-learn.org/stable/auto_examples/neighbors/plot_nca_classification.html.
- [3] *People Count User's Guide*. Dallas, Spojené státy americké: Texas Instruments, září 2019. Verze SDK: MMWAVE-SDK 03_03_00_X.
- [4] *Traffic Monitoring Object Detection and Tracking Reference Design Using Single-Chip mmWave Radar Sensor* [online]. Dallas, Spojené státy americké: Texas Instruments, květen 2019. Dostupné z: <http://www.ti.com/lit/ug/tidud31b/tidud31b.pdf?ts=1591045306651>.
- [5] *MmWave radar sensors- What is mmWave* [online]. Dallas, Spojené státy americké: Texas Instruments, 2020 [cit. 2020-05-25]. Dostupné z: <http://www.ti.com/sensors/mmwave-radar/what-is-mmwave.html>.
- [6] AGGARWAL, N. a KARL, W. Line detection in images through regularized Hough transform. In: *IEEE Transactions on Image Processing*. Rochester, Spojené státy americké: IEEE, Leden 2006, s. 582–591. ISSN 1941-0042.
- [7] ASAWA, C. *CS231n: Convolutional Neural Networks for Visual Recognition* [online]. Redwood City, Spojené státy americké: Stanford University Press, 2019 [cit. 2019-11-26]. Dostupné z: <http://cs231n.github.io/>.
- [8] DAVID ALDAVERT, M. R. *Manuscript Text Line Detection and Segmentation using Second-Order Derivatives* [online]. Barcelona, Španělsko: Universitat Autònoma de Barcelona, 2018 [cit. 2020-01-07]. Dostupné z: http://www.cvc.uab.cat/people/aldavert/documents/conferences/das2018_line_segmentation.pdf.
- [9] G. EL NATOUR, F. R. F. R. P. B. Radar and vision sensors calibration for outdoor 3D reconstruction. In: *Proceedings - IEEE International Conference on Robotics and Automation*. Spojené státy americké: IEEE, červen 2015, s. 2084–2089.
- [10] ISABELLE GUYON, A. E. An Introduction to Variable and Feature Selection. In: *Journal of Machine Learning*. první edice. Brookline, Spojené státy americké: Microtome Publishing, Březen 2003 [cit. 2020-05-05]. ISSN 1532-4435.

- [11] J. HEIKKILA, O. S. A four-step camera calibration procedure with implicit image correction. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Juan, Spojené státy americké: IEEE, 1997, s. 1106–1112. ISBN 0-8186-7822-4.
- [12] J. R. R. UIJLINGS, T. G. a. A. W. M. S. *Selective Search for Object Recognition* [online]. Berlin, Heidelberg: Springer, 2013 [cit. 2020-01-08]. Dostupné z: <https://link.springer.com/article/10.1007/s11263-013-0620-5>.
- [13] LIU, Y. H. *Feature Extraction and Image Recongition with Convolution Neural Networks* [online]. Bristol, Velká Británie: IOP Publishing, 2018 [cit. 2020-05-16].
- [14] LUYANG JING, M. Z. P. W. *An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox*. Basel, Švýcarsko: MDPI, únor 2017.
- [15] MICHAL ŠPANĚL, V. B. *Obrazové segmentační techniky* [online]. Brno, Česká republika: Nakladatelství Vysokého učení technického v Brně VUTIUM, 2006 [cit. 2020-01-07]. Dostupné z: <http://www.fit.vutbr.cz/~spanel/segmentace/>.
- [16] NITISH SRIVASTAVA, A. K. I. S. R. S. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. In: *Journal of Machine Learning Research*. Brookline, Spojené státy americké: Microtome Publishing, 2014, s. 1929–1958. Dostupné z: <http://jmlr.org/papers/v15/srivastava14a.html>.
- [17] PEDRO F. FELZENSZWALB, D. P. H. *Efficient Graph-Based Image Segmentation* [online]. Berlin, Heidelberg: Springer, 2004 [cit. 2020-01-08]. Dostupné z: <http://cs.brown.edu/people/pfelzens/papers/seg-ijcv.pdf>.
- [18] PETRŽELA, L. *Aplikace senzorů pracujících v oblasti milimetrových vln v zabezpečovací technice*. Brno, 2019. Disertační práce. Vysoké učení technické v Brně. Vedoucí práce Soběslav Valach.
- [19] R. O. DUDA, P. E. H. a. D. G. S. *Pattern Recognition*. 2. vyd. New York, Spojené státy americké: John Wiley & Sons, 2000 [cit. 2020-03-15]. ISBN 04-710-5669-3.
- [20] RENYUAN ZHANG, S. C. Extending Reliability of mmWave Radar Tracking and Detection via Fusion with Camera. In: *IEEE Access*. 1. vyd. Spojené státy americké: IEEE, Záhř 2019.
- [21] SERGEY IOFFE, C. S. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Ithaca, Spojené státy americké: Cornell University Press, 2015.
- [22] TAMMINA, S. *Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images* [online]. New York, Spojené státy americké: IJSRP (International Journal of Scientific and Research Publications), říjen 2019 [cit. 2020-05-16]. Dostupné z: <http://www.ijsrp.org/research-paper-1019.php?rp=P949194>.
- [23] TAO WANG, J. X. Z. M. Integrating Millimeter Wave Radar with a Monocular Vision Sensor for On-Road Obstacle Detection Applications. In: *Sensors*. Basel, Švýcarsko: MDPI, 2011 [cit. 2020-05-11]. ISSN 1424-8220.

- [24] TIAN, D. A review on image feature extraction and representation techniques. In: *International Journal of Multimedia and Ubiquitous Engineering*. Jižní Korea: Science and Engineering Research Support Society, Leden 2013. ISSN 1975-0080.
- [25] VLADIMÍR MARŠÍK, J. L. a. k. *Umělá inteligence*. 1. vyd. Praha, Česká republika: Academia Praha, 1993 [cit. 2020-04-07]. ISBN 80-200-0496-3.
- [26] Y. LECUN, J. S. D. D. H. R. E. H. W. H. L. D. J. *Backpropagation applied to handwritten zip code recognition*. Frechen, Německo: MITP, 1989 [cit. 2020-05-16].
- [27] YANG, C. *How Many Pixels Do We Need to See Things?* [online]. Berlin, Heidelberg: Springer, 2003 [cit. 2020-05-14]. Dostupné z: https://link.springer.com/content/pdf/10.1007/3-540-44863-2_105.pdf.
- [28] YANN LECUN, Y. B. P. H. *Gradien-Based Learning Applied to Document Recognition* [online]. New York, Spojené státy americké: IEEE, listopad 1998 [cit. 2020-05-16]. Dostupné z: <http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>.
- [29] ZHAO, Z.-Q., ZHENG, P., XU, S. tao a WU, X. Object Detection with Deep Learning: A Review. In: *IEEE Transactions on Neural Networks and Learning Systems*. Kingston, Spojené státy americké: IEEE, 2019. ISSN 2162-2388.

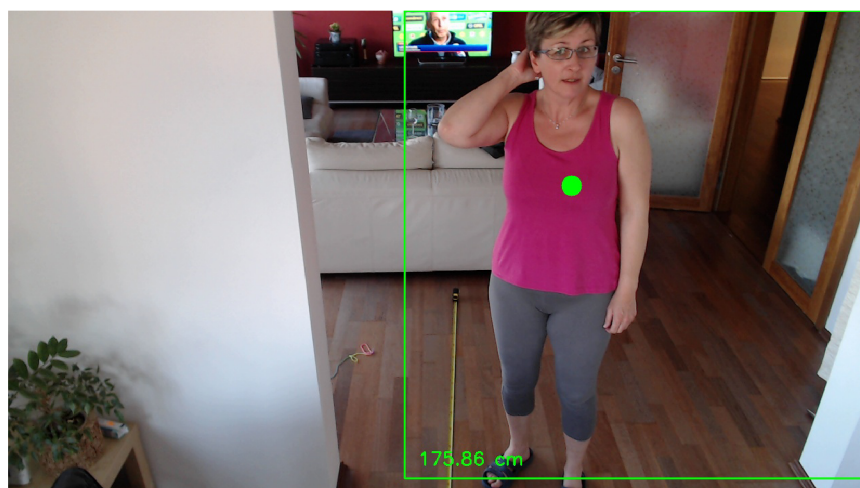
Seznam příloh

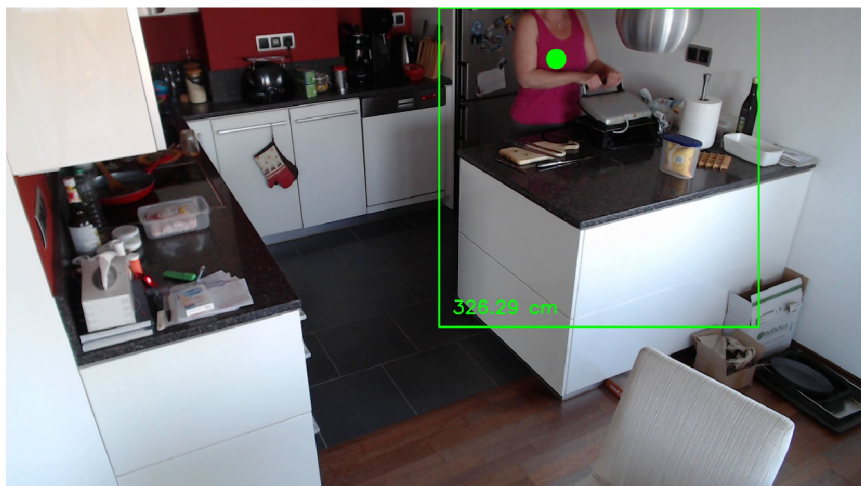
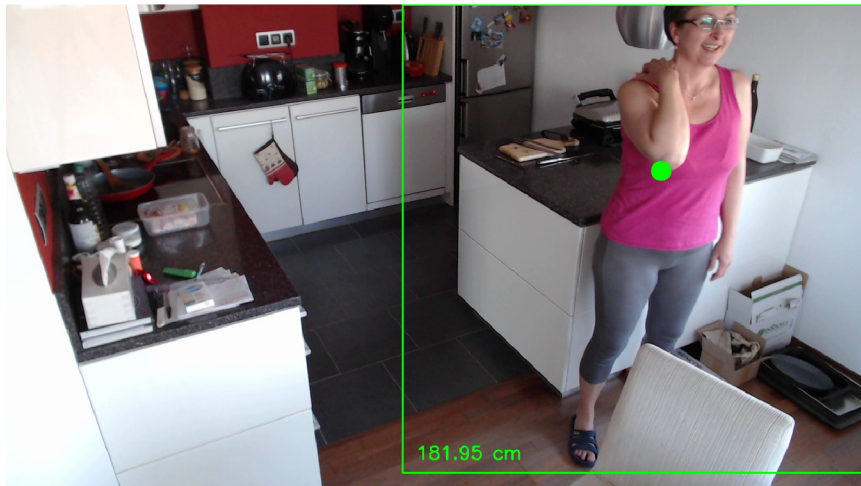
A	Výsledky testovacích vstupů	48
A.1	Příklady výsledků testovacích vstupů detekce	48
A.2	Příklady výsledků test. vstupů ve zhoršených světelných podmínkách	52
B	Příklady trénovacích vstupů konvoluční neuronové sítě	54
B.1	Příklady trénovacích vstupů osob	54
B.2	Příklady trénovacích vstupů prostředí	55

Příloha A

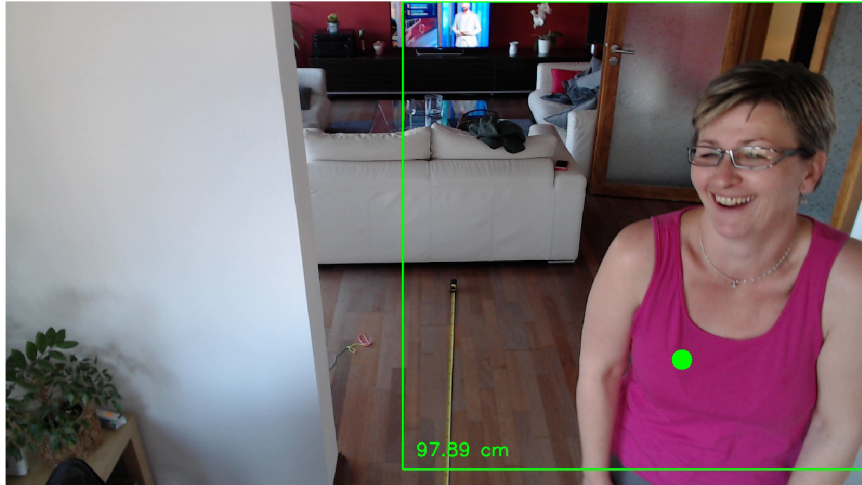
Výsledky testovacích vstupů

A.1 Příklady výsledků testovacích vstupů detekce

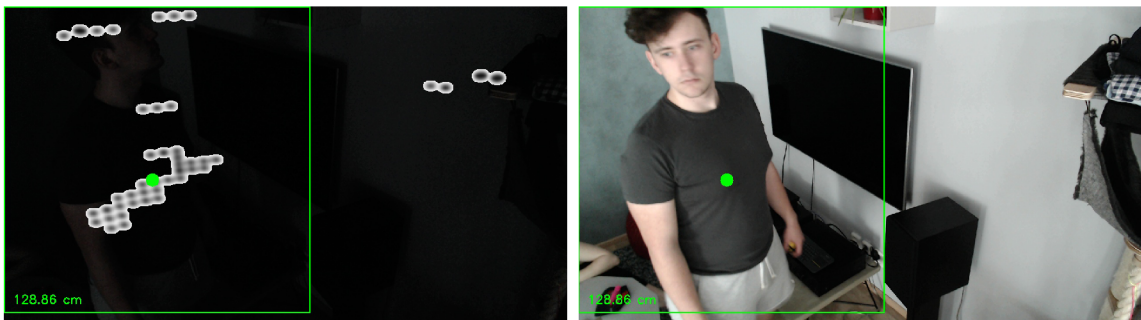
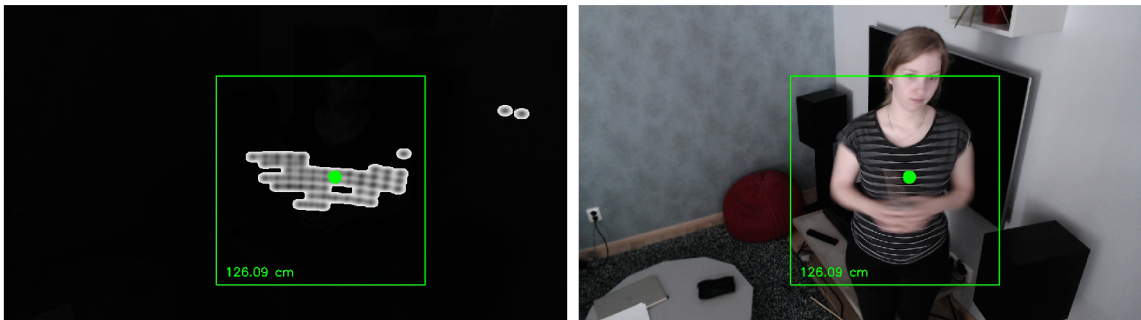
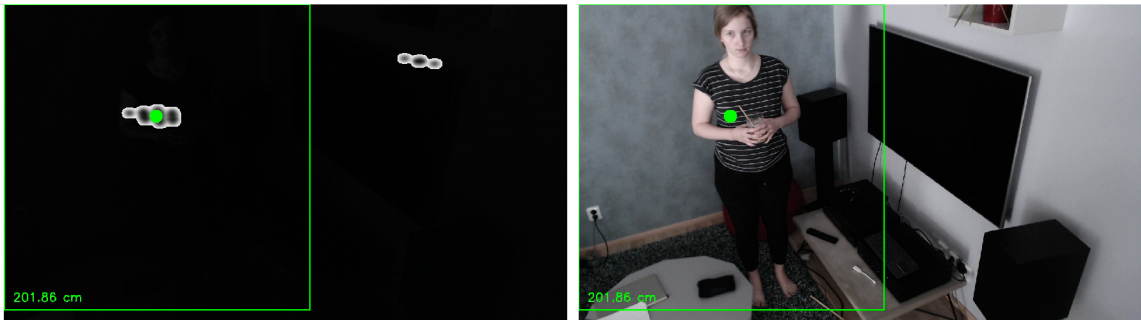








A.2 Příklady výsledků test. vstupů ve zhoršených světelných podmínkách

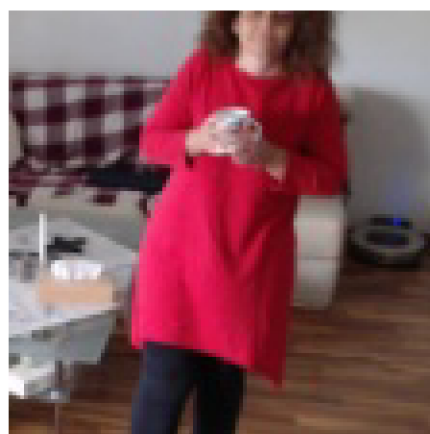
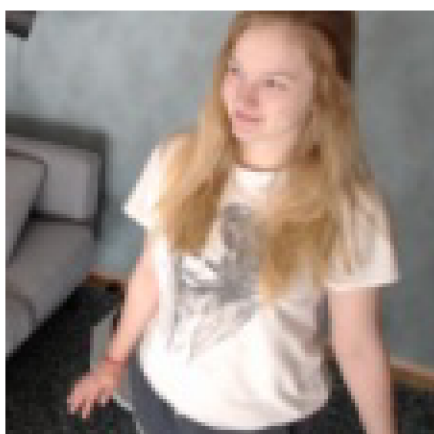
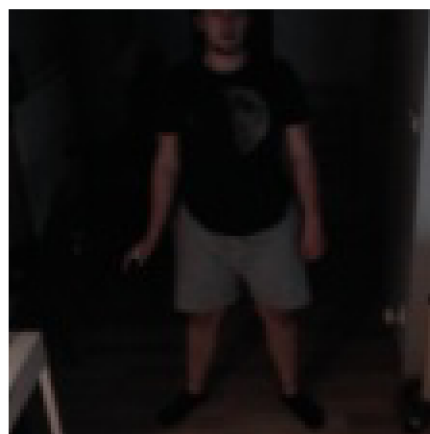
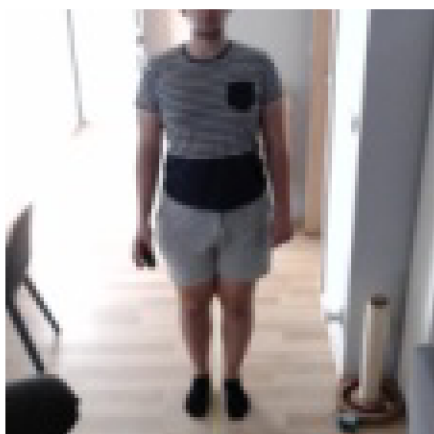


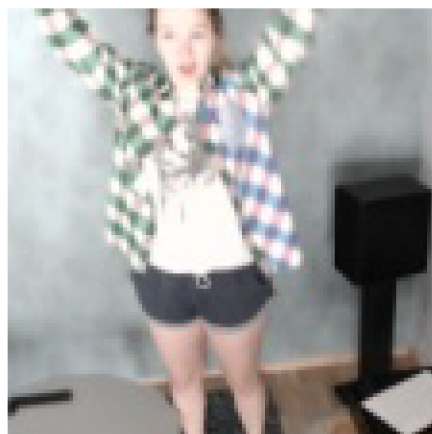
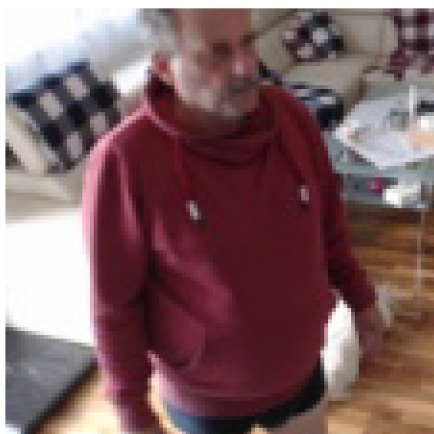


Příloha B

Příklady trénovacích vstupů konvoluční neuronové sítě

B.1 Příklady trénovacích vstupů osob





B.2 Příklady trénovacích vstupů prostředí

