



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

DEPARTMENT OF INTELLIGENT SYSTEMS

**ROZPOZNÁVÁNÍ OSOB NA ZÁKLADĚ ANTROPOMET-
RICKÝCH PROPORCÍ**

PERSON RECOGNITION BASED ON ANTHROPOMETRIC PROPORTIONS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

SAMUEL ŠIMÚN

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. TOMÁŠ GOLDMANN

BRNO 2024

Zadání bakalářské práce



156902

Ústav: Ústav inteligentních systémů (UITS)
Student: **Šimún Samuel**
Program: Informační technologie
Název: **Rozpoznávání osob na základě antropometrických proporcí**
Kategorie: Umělá inteligence
Akademický rok: 2023/24

Zadání:

1. Seznamte se s možnostmi detekce a rozpoznáváním osob v obraze.
2. Nastudujte možnosti využití antropometrických proporcí těla pro rozpoznání osob v obraze.
3. Navrhněte neuronovou síť pro rozpoznávání osob na základě antropometrických proporcí. Neuronová síť bude využívat data získaná pomocí existující neuronové sítě pro odhad postoj osob.
4. Navrženou neuronovou síť implementujte v programovacím jazyce Python a vytvořte k ní jednoduchou aplikaci.
5. Proveďte experimenty zaměřené na vyhodnocení přesnosti rozpoznávání osob v implementovaném řešení.

Literatura:

- WANG, Jinbao, et al. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*, 2021, 210: 103225.
- ZHENG, Ce, et al. Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*, 2023, 56.1: 1-37.

Při obhajobě semestrální části projektu je požadováno:
Body 1 a 2.

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

Vedoucí práce: **Goldmann Tomáš, Ing.**
Vedoucí ústavu: Hanáček Petr, doc. Dr. Ing.
Datum zadání: 1.11.2023
Termín pro odevzdání: 9.5.2024
Datum schválení: 6.11.2023

Abstrakt

Táto bakalárska práca sa zameriava na možnosti využitia biometrie a strojového učenia. Podrobne rozoberá antropometrické charakteristiky, ich vývoj a potenciál využitia týchto charakteristik pre identifikáciu osoby. Analyzuje možnosti strojového učenia a aké technológie existujú pre extrakciu pózy osoby z obrazu pre výstup pózy v druhej dimenzii a v tretej dimenzii. Taktiež práca popisuje a implementuje konkrétne metódy ktoré využívajú antropometrické vlastnosti pre identifikáciu osôb.

Abstract

This bachelor thesis focuses on the potential use of biometrics and machine learning. It discusses in detail anthropometric characteristics, their evolution and the potential of using these characteristics for person identification. It analyzes the potential uses of machine learning and what technologies exist for extracting a person's pose from an image for outputting the pose in the second dimension and in the third dimension. Also, the thesis describes and implements specific methods that use anthropometric features for person identification.

Klíčové slová

Biometria, Antropometrické vlastnosti, Strojové učenie, Odhad 2D pózy z obrazu, Odhad 3D pózy z obrazu

Keywords

Biometrics, Anthropometric features, Machine learning, 2D pose estimation, 3D pose estimation

Citácia

ŠIMŮN, Samuel. *Rozpoznávání osob na základě antropometrických proporcí*. Brno, 2024. Bakalárska práca. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Tomáš Goldmann

Rozpoznávání osob na základě antropometrických proporcí

Prehlásenie

Prehlasujem, že som túto bakalársku prácu vypracoval samostatne pod vedením pána inžiniera Goldmanna. Uviedol som všetky literárne pramene, publikácie a ďalšie zdroje, z ktorých som čerpal.

.....
Samuel Šimún
7. mája 2024

Podakovanie

Rád by som vyjadril svoju vďaku pánovi inžinierovi Tomášovi Goldmannovi za vedenie mojej bakalárskej práce a za poskytnutie odborných konzultácií pri tvorbe tejto práce.

Obsah

1	Úvod	5
2	Biometria a antropometrické vlastnosti tela	6
2.1	Biometria	6
2.1.1	Behaviorálna biometria	6
2.1.2	Fyzická biometria	7
2.2	Biometrické systémy	7
2.2.1	Hodnotenie chybovosti biometrických systémov	8
2.2.2	Spracovanie oblasti tváre	10
2.3	Antropometria	11
2.3.1	Prehľad antropometrických vlastností	11
2.3.2	Vývin antropometrických vlastností	13
2.3.3	Vývin u dospelých jedincoch	14
2.3.4	Význam antropometrických vlastností	14
2.4	Strojové učenie a extrakcia antropometrických vlastností	15
3	Extrakcia antropometrických rysov postavy	16
3.1	Strojové učenie	16
3.1.1	Typy strojového učenia	16
3.1.2	Neurónové siete	18
3.1.3	Konvolučné neurónové siete	18
3.2	Odhad pózy osoby na obraze	19
3.2.1	Detekcia osoby na obraze	19
3.2.2	Detekcia kľúčových bodov	20
3.2.3	YOLO Pose	21
3.3	Detekcia pózy osoby v 3D	23
3.3.1	Výzvy pre detekciu pózy v 3D	23
3.3.2	Modelovanie pózy	24
3.3.3	Prístupy k detekcii pózy	25
3.3.4	Metriky pre výpočet chyby	25
4	Príprava dát a rozpoznávanie osôb s využitím strojového učenia	27
4.1	Návrh riešenia	27
4.2	Dátové sady	27
4.2.1	Large-Scale Spatio-Temporal Person Re-identification Dataset (<i>LaST</i>)	28
4.2.2	Player Reidentification Challenge Dataset	28
4.3	Spracovávanie obrazového datasetu	29
4.3.1	Model pre estimáciu 3D pózy	29

4.3.2	Spracovanie výsledkov	29
4.3.3	Analýza výsledných dát	31
4.4	Trénovanie Machine Learning modelu	33
4.4.1	Trénovanie neurónovej siete	33
4.4.2	Trénovanie ML modelu pomocou nástroja AutoML	36
5	Analytické a vizuálne rozpoznávanie osôb	37
5.1	Rozpoznávanie osôb na základe podobnosti vektorov	37
5.1.1	Návrh riešenia	37
5.1.2	Rozpoznávanie osôb podľa vzdialenosti vektorov	38
5.1.3	Rozpoznávanie osôb s použitím transformácie pózy	39
5.1.4	Rozpoznávanie osôb podľa chyby pózy	42
5.2	Klasifikácia osôb na základe vizuálnej podobnosti	43
5.2.1	Výber detekčného modelu	44
5.2.2	Metrika pre porovnanie tried	44
5.2.3	Výber embedding modelu	44
5.2.4	Návrh systému	45
5.2.5	Vyhodnotenie metódy	45
5.3	Súhrn uskutočnených experimentov	46
5.3.1	Súhrnné testovanie metód	47
6	Záver	49
	Literatúra	51
A	Obsah pamäťového média	55

Zoznam obrázkov

2.1	Príklad záberu kamery kde nie je možné spoľahlivo extrahovať tvár. <i>Prevzaté a upravené z datasetu LAsT</i> [27].	8
2.2	Príklad záberu kamery kde je možné spoľahlivo extrahovať tvár osoby. <i>Prevzaté a upravené z datasetu LAsT</i> [27].	8
2.3	Oblasti a konkrétne veličiny ľudského tela sledované biometrickými systémami. <i>Prevzaté z</i> [22].	9
2.4	Graf závislosti pravdepodobnosti od rozdelenia skóre pri Autentickej distribúcii a Distribúcii podvodníka. <i>Prevzaté z</i> [5].	10
2.5	Prehľad základných antropometrických vlastností človeka. <i>Prevzaté z</i> [16].	12
3.1	Rozdelenie ML modelov a ich využitie. <i>Prevzaté z</i> [21].	17
3.2	Yolo pose architektúra ktorá je založená na YOLOv5. <i>Prevzaté z</i> [17].	21
3.3	Príklad SMPL modelu vyjadrujúceho ľudskú pózu. Biele body na obrázku vyjadrujú pred definované body. <i>Prevzaté z</i> [32].	24
4.1	Návrh riešenia s využitím strojového učenia.	28
4.2	Príklad inferencie modelu MeTRAbs. Vľavo osoba a jej póza v 2D. Vpravo póza osoby v 3D priestore.	30
4.3	Neurónová sieť tvorená podľa návrhu 4.4.1 so 25 vstupnými triedami.	33
4.4	Graf označujúci presnosť modelu počas učenia na datasete LaST	35
4.5	Graf označujúci hodnotu stratovej funkcie počas učenia na datasete LaST	35
4.6	Graf označujúci presnosť modelu počas učenia na datasete Player Reidentification Challenge	35
4.7	Graf označujúci hodnotu stratovej funkcie počas učenia na datasete Player Reidentification Challenge	35
4.8	Dôležitosť/váha hodnoty atribútu na rozhodnutie modelu.	36
5.1	Návrh riešenia s využitím analytického rozpoznávania osôb.	38
5.2	Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu podobnosti vektorov.	39
5.3	Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu podobnosti vektorov.	39
5.4	Referenčná kostra vytvorená nad skráteným datasetom Player Reidentification Challenge.	40
5.5	Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu s transformáciou pózy.	41
5.6	Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu s transformáciou pózy.	41

5.7	Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu chyby pózy.	43
5.8	Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu chyby pózy.	43
5.9	Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu vizuálnej podobnosti.	45
5.10	Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu vizuálnej podobnosti.	45

Kapitola 1

Úvod

Medzi dôležité aspekty dnešnej doby patrí potreba rýchleho a spoľahlivého rozpoznávania osôb. Existujúce riešenia, známe aj z reálneho sveta, sa sústreďujú na rozpoznávanie osôb na základe častí tela ako tvár, odtlačky prstov alebo dúhovky. Tieto časti tela sú známe tým, že obsahujú ľahko extrahovateľné atribúty, ktoré sú v rámci populácie dostatočne unikátne na to, aby bolo možné na základe týchto atribútov určiť identitu osoby.

Aj keď sú tieto vlastnosti relatívne jednoduché na extrahovanie, je pre ich získanie často nutné aby daná časť tela bola celá vystavená snímaču. To v rôznych situáciách znižuje počet vzorkov ktoré je možné snímaním získať a pomocou nich následne rozpoznať osobu.

Ako riešenie môže byť rozpoznávanie osôb na základe vlastností ktoré sú viditeľné na kamerách najčastejšie a najdlhšie - časti osoby. Každá časť tela osoby nesie v sebe informácie, ktoré môžu byť na rozpoznávanie využité. Ide napríklad o dĺžku ruky, veľkosť ramien, dĺžku stehennej kosti a podobne.

Pre účely získavania týchto vlastností je možné využiť už existujúce technológie, ktoré dokážu z obrazu predpovedať pózu v 3D priestore. Na základe výstupnej kostry sa ďalej ponúkajú rôzne prístupy ako docieľiť k efektívnemu systému, ktorý dokáže vstupný snímok osoby správne prideliť ku záznamu v znalostnej databáze a tým danú osobu identifikovať.

Cielom danej práce je celý proces analyzovať a realizovať. Je nutné zistiť či antropometrické vlastnosti je možné využiť na rozpoznávanie osôb a do akej miery sú vhodné na vytvorenie biometrického systému. Ďalej je potrebné zistiť, ako je presné získavanie antropometrických vlastností modelom a či je presnosť dostatočne vysoká na to, aby bola identifikácia osoby vôbec možná.

Kapitola 2

Biometria a antropometrické vlastnosti tela

Táto kapitola má za cieľ čitateľa zoznámiť s existujúcimi možnosťami a prostriedkami pomocou ktorých je možné na snímkoch detekovať a rozpoznávať osoby. Rozpoznávanie je uskutočňované na základe určitých charakteristík ktoré danú osobu definuje. Tieto vlastnosti však musia spĺňať účel jedinečnosti (vlastnosť musí byť rozlíšiteľná medzi väčšou časťou populácie na to aby dávalo zmysel segmentovať a kvalitnejšie rozpoznávať osoby), stálosti, zberateľnosti (požadovanú vlastnosť je nutné nejakým spôsobom zachytiť a digitalizovať). V nasledujúcich kapitolách sa čitateľ môže dozvedieť aké prostriedky pre rozpoznávanie môžu byť použité a aké vlastnosti sa v existujúcich systémoch bežne vyžívajú.

2.1 Biometria

Slovo biometria (z grec. 'bios' - život, 'metrikos' - meranie) označuje odvetvie vedy, ktorej cieľom je štatisticky analyzovať biologické charakteristiky. Biometria sa taktiež zaoberá možnosťami a spôsobmi, ako pomocou daných biologických charakteristík identifikovať danú osobu. Tieto charakteristiky môžeme vo všeobecnosti rozdeliť do dvoch podskupín : **behaviorálne** a **fyzické**, ktorých využitie je možné sledovať aj na obrázku 2.3 [15, 6].

2.1.1 Behaviorálna biometria

Táto podskupina biometrie sa sústreďuje na merania a údaje ktoré boli získané od užívateľa prostredníctvom vykonávania určitej činnosti. Vo všeobecnosti ide o každú vlastnosť osoby ktorú vykonáva prirodzene, vychádzajúcich z jeho správania. Medzi tieto vlastnosti je možné zaradiť dĺžku pri chôdzi, gestá, podpis alebo aj hlas [6].

Tieto charakteristiky sú oproti fyzickým náročnejšie na extrakciu a zachytenie. Taktiež sú tieto vlastnosti ovplyvnené vonkajšími vplyvmi ako starnutie, ktoré môže hlavne u dospievajúcich jedincov zmeniť tóninu hlasu alebo vplyvom rastu zvýšiť dĺžku kroku (popísané v sekcii 2.3.2) [15].

Ich využitie sa však vplyvom trendov v biometrii stáva čoraz častejšie. Ide hlavne o systémy ktoré sa prostredníctvom umelej inteligencie priebežne učia na správaní užívateľa. Týmto spôsobom sa za určitý čas vedia naučiť dostatočné množstvo charakteristík danej osoby pomocou ktorých je následne možné kvalitnejšie a efektívnejšie identifikovať danú osobu [15].

2.1.2 Fyzická biometria

Fyzická biometria sa zameriava na meranie určitej časti ľudského tela. Sústredí sa hlavne na časti tela, ktoré sú čo najviac odlišiteľné od ostatných osôb a svojim spôsobom danú osobu charakterizujú. Medzi časti tela ktoré sú najčastejšie používané na identifikáciu patrí: odtlačky prstov, tvár, uši, rohovky oka a ďalšie iné [6].

Pre zachytenie týchto charakteristík je možné využiť rôzne kamery a snímače. Kvalita následného rozpoznávania však závisí na kvalite daného snímača alebo kamery a taktiež aj na kvalite odobraného vzorku (napr. polohovanie prstu na snímač pre získanie odtlačku) [15].

Využitie týchto vlastností je aktuálne najviac typickým spôsobom ako biometricky identifikovať osobu. Tieto vlastnosti sú dostatočne unikátne, čo zabezpečuje systém pred existenciou dvoch identít ktoré by mali zhodnú biometrickú charakteristiku. Avšak až kvalita snímača a voľba metódy porovnávania vie v reálnom systéme zabezpečiť rýchlu a spoľahlivú identifikáciu [6].

Ako príklad využitia fyzickej biometrie môžeme uviesť rozpoznávanie osôb na základe odtlačkov prsta. To predstavuje jeden z najstarších biometrických spôsobov pre rozpoznávanie. Základnými atribútmi odtlačkov prstov sú hrebienky a dolinky končekov prstov. Tie tvoria unikátny vzor pomocou ktorého je možné správne určiť identitu osoby [1].

Pri porovnávaní odtlačkov sú najčastejšie používané dve metódy: **minutiálne¹ porovnávanie** ktorá analyzuje a porovnáva odtlačky na základe minútii a **makroskopický prístup** ktorý sa zameriava na na priebeh hrebeňov, teda na ich oblúky a vršky [1].

Taktiež veľmi dôležitá časť rozpoznávania je práve zachycovanie a digitalizácia odtlačkov prstov ktorá sa taktiež rozdeľuje do dvoch kategórii. A to do **priamej** a **nepriamej**. Nepriame zachytávanie odtlačkov prstov spočíva v nanášaní atramentu alebo inej značiacej tekutiny na prst a následný otláčok na papier. Z tohto papiera sa následne vytvorí snímok pomocou ktorého sa následne typicky identifikuje osoba. Tento postup nie je najpresnejší a to hlavne z hľadiska toho že atrament alebo papier nemusí vždy správne zachytiť odtlačok alebo môže vznikáť šum na skene papiera. Priame zachytávanie spočíva v použití špeciálnych skenerov ktoré sú vytvorené pre tento účel. Spočíva v tom že osoba priloží prst na skener a skener vráti digitalizovaný odtlačok prsta ktorý je značne presnejší na rozdiel od nepriameho zachytávania [1, 3].

2.2 Biometrické systémy

Biometrické systémy môžeme označovať ako systémy pre rozpoznávanie vzorov. Tento systém rozpoznáva osobu na základe vektoru vlastností (*feature vector*) odvodeného od fyziologickej alebo behaviorálnej charakteristiky ktorou konkrétna osoba disponuje. Ich účelom je hlavne potvrdiť alebo vyvrátiť skutočnosť že ide jedinca ktorého totožnosťou sa jedinec identifikuje [5].

Tento vektor príznakov najčastejšie býva uložený do databázy v ktorej slúži na následnú verifikáciu osoby, ktorá sa chce na základe svojho vektoru vlastností identifikovať. Biometrický systém dokáže fungovať v dvoch režimoch. A to v **identifikačnom** režime kedy porovnáваме šablónu vektoru vlastností so všetkými vektormi vlastností ktoré sa nachá-

¹Špecifické body, kde sa mení vzor hrebienkov na odtlačku prsta

dzajú v databáze alebo **overovacom** režime kedy porovnávame šablónu vektoru vlastností iba s vektormi ktoré zodpovedajú identite deklarovanej osobe. Z popisu daných problémov teda vyplýva že ide o dva postupy ktoré je vhodné riešiť samostatne [5].

Vlastnosti ktoré sa využívajú pre identifikáciu a verifikáciu osôb je viacero. Ako bolo spomenuté v úvode tejto sekcie 2 a v pod-sekcii úvode do biometrie 2.1, biometrických vlastností môže byť teoreticky každá fyziologická alebo behaviorálna vlastnosť osoby ktorá spĺňa vlastnosti ako jedinečnosť, stálosť, zberateľnosť alebo prístupnosť. Príklady týchto vlastností je možné pozorovať na obrázku 2.3. V nasledujúcich podkapitolách sa text bude zaoberať typickými biometrickými vlastnosťami ktoré sú bežne používané v praxi [22].



Obr. 2.1: Príklad záberu kamery kde nie je možné spoľahlivo extrahovať tvár. *Prevzaté a upravené z datasetu LAsT [27].*

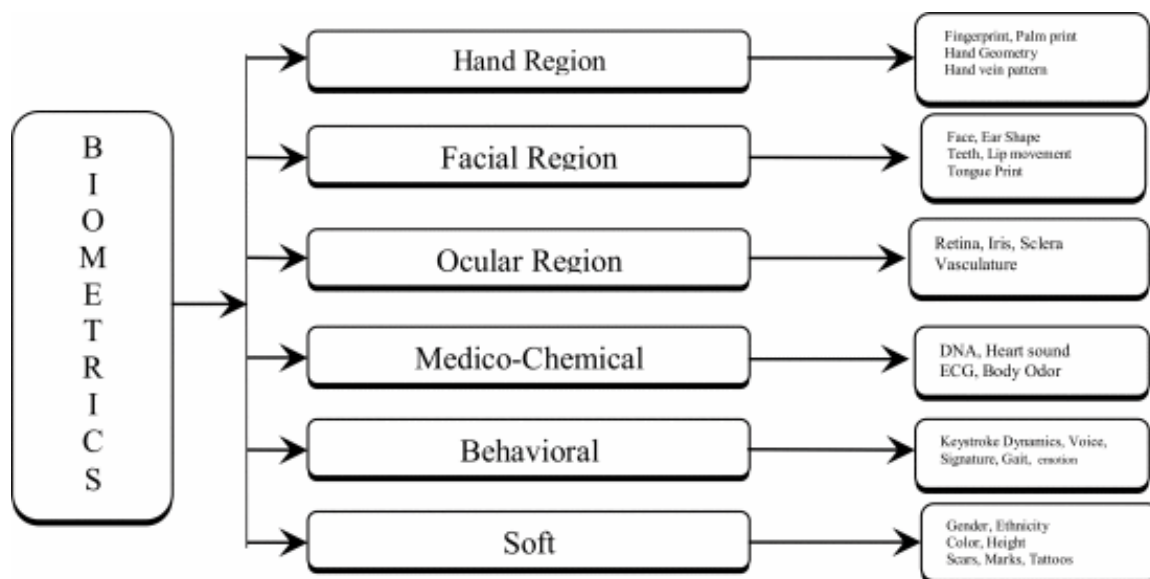


Obr. 2.2: Príklad záberu kamery kde je možné spoľahlivo extrahovať tvár osoby. *Prevzaté a upravené z datasetu LAsT [27].*

V rámci záberov z bezpečnostných kamier môžu nastať dve situácie: na zábere je kvalitne zachytená celá tvár ako je možné sledovať na obrázku 2.2, čo umožňuje identifikáciu osoby na základe tváre (popísané v pod-sekcii 2.2.2) alebo môže nastať že osoba je ku kamere otočená takým uhlom, že nie je spoľahlivo možné extrahovať rysy tváre z osoby ako napríklad na obrázku 2.1. Na kamere je však možné vidieť ostatné časti tela na základe ktorých je možná identifikácia pomocou antropometrických vlastností (popísané v sekcii 2.3).

2.2.1 Hodnotenie chybovosti biometrických systémov

Rôzne vonkajšie faktory ako zmena vonkajšieho prostredia, šum, deformácia, zlé umiestnenie na senzore, šum alebo zlá iterácia zo strany klienta vedie k tomu, že nie je možné aby sa dve vzorky toho istého objektu (osoby) a jej biometrickej charakteristiky zhodovali 1:1. Z tohoto dôvodu sa najčastejšie preto používa hodnota, ktorá charakterizuje presnosť zhody (*confidence score*) medzi dvoma vzorkami a tým pádom kvantifikuje podobnosť medzi vstupným a databázovým vzorom. Čím je presnosť zhody (*confidence score*) vyššie tým



Obr. 2.3: Oblasti a konkrétne veličiny ľudského tela sledované biometrickými systémami. Prevzaté z [22].

si je systém istejší že dané porovnávané dve vzorky sú zhodné a tým pádom patria jednému človeku [5].

Pre rozhodnutie či je osoba identifikovaná pomocou jej vzorku z databázy sa využíva prah akceptovateľnosti, ktorý značí do akej miery musia byť vzorky zhodné aby bolo možné prehlásiť že dané dve vzorky patria tej istej osobe. Naopak pokiaľ by bolo skóre nižšie ako prahová hodnota, vzorky patria dvom rôznym osobám.

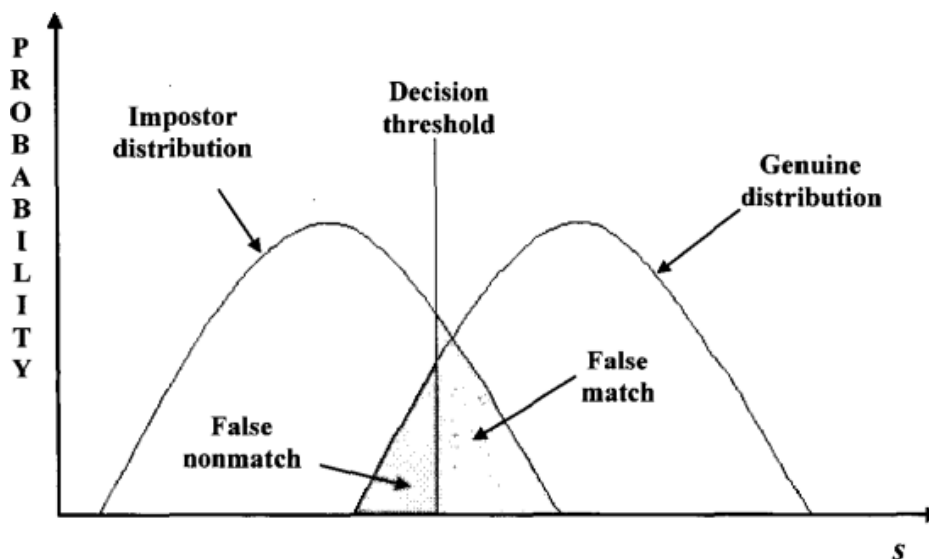
Pre vyjadrenie presnosti a výkonnosti biometrických systémov sa používajú dva koncepty: **Autentická distribúcia** (*Genuine Distribution*) a **Distribúcia podvodníka** (*Impostor Distribution*). Autentická distribúcia odkazuje na distribúciu presnosti zhôd, v ktorej sa porovnávajú páry biometrických vzoriek jednej danej osoby. Naopak distribúcia podvodníka reprezentuje distribúciu presnosti zhôd, ktorej sa porovnávajú páry biometrických vzoriek dvoch nerovnakých osôb. Obe distribúcie ktoré je možné sledovať na obrázku 2.4 značia príklad reálneho systému kde sa dve dané distribúcie prelínajú [1].

V kontexte merania presnosti biometrického systému využívame dva pojmy [5]:

- **FNMR** Miera falošnej zhody (*false nonmatch rate*) - miera ktorá označuje situáciu ku ktorej dochádza, keď dve vzorky rôznych osôb sú označené za zhodné. Výpočet FNMR je nasledujúci:

$$\text{FNMR} = \frac{\text{Počet falošných nezhôd (False nonmatch)}}{\text{Celkový počet pokusov o zhodu}} \quad (2.1)$$

- **FMR** Miera falošnej nezahody (*false match rate*) - miera ktorá označuje situáciu ku ktorej dochádza, keď dve vzorky rovnakých osôb sú označené za rôzne vzorky (nejde o tú istú osobu). Výpočet FMR je nasledujúci:



Obr. 2.4: Graf závislosti pravdepodobnosti od rozdelenia skóre pri Autentickej distribúcii a Distribúcii podvodníka. *Prevzaté z [5].*

$$\text{FMR} = \frac{\text{Počet falošných zhôd (False match)}}{\text{Celkový počet pokusov o nezhodu}} \quad (2.2)$$

Svojím spôsobom sú FNMR a FMR funkcie prahovej hodnoty. Pokiaľ znížime prah istoty znamená to že sa zvýši FMR ale na druhú stranu pokiaľ prah istoty zvýšime, rovnako tak aj zvýšime FNMR čo vedie že systém je bezpečnejší ale ťažšie použiteľný [5].

2.2.2 Spracovanie oblasti tváre

Rozpoznávanie osôb na základe tváre je jednou z hlavných spôsobov rozpoznávania jedincov a to hlavne z dôvodu toho že základnou ľudskou vlastnosťou je práve rozpoznávať a identifikovať jedincov na základe ich tváre. Algoritmy na tvári si všímajú dôležité vlastnosti pre rozpoznávanie ako vzdialenosť medzi očami, veľkosť nosu, čela a podobne. Na základe týchto vlastností ďalej algoritmus analyzuje a pokúša sa o zhodu vzorov ktoré má k dispozícii. Z toho plynie že základný proces pri identifikácii osoby je [9, 22]:

- **Detekovanie a lokalizovanie tváre v obraze** - Proces ktorý zahŕňa lokalizovanie existujúcich tvári na obraze. Najčastejšie sa pre to využívajú Konvolučné Neurónové siete (popísané v sekcii 3.1.3) ktoré dokážu na obraze nájsť tváre.
- **Extrahovanie znakov tváre** - Tento proces obnáša získavanie vlastností tváre po detekovaní v predchádzajúcom kroku. Algoritmus hľadá vlastnosti ktoré boli popísané v úvode tejto sekcii 2.2 (vzdialenosť medzi očami, pozícia nosu, atď).
- **Porovnávanie vlastností** - Po získaní vlastností z tváre je nutné aby bola daná osoba identifikovaná alebo verifikovaná. Tento proces zahŕňa porovnávanie vlastností tváre s už existujúcimi vektormi v databázy.

Vlastnosti ktoré sa sledujú pri extrahovaní znakov tváre sa líšia podľa toho aký typ algoritmu sa pri rozpoznávaní tváre použije. Medzi najznámejšie algoritmy patria: *Eigen Faces*, *Fisher faces*, *Scale Invariant Feature Transform* [9, 22].

Využitie rozpoznávania tváre v praxi

Veľkou výhodou rozpoznávania osôb na základe tváre je rozpoznávanie bez nutnosti zásahu človeka. To znamená že využitie že rozpoznávaná osoba nemusí priamo iterovať so žiadnym objektom, ale stačí aby osoba bola natočená tvárou ku snímaču (napr. kamere). Veľké využitie si tento prístup vie nájsť vo verejných priestranstvách ako nákupné centrá alebo letiská [9].

Medzi hlavné problémy ktoré bránia správne rozpoznávaniu patria hlavne problémy s osvetlením, zmenená póza na snímku alebo make-up u osoby. Práve kolísanie osvetlenia a kolísanie polohy tváre majú najväčší dopad na správne rozpoznávanie. V praxi môžu nastať situácie, kedy sa na obraze tváre vyskytne tieň alebo sa zmení gradient kontrastu. Práve tieto faktory môžu do značnej miery zhoršiť presnosť predpovede, čo môže viesť u systémov nasadených v praxi značné problémy s identifikáciou [9].

2.3 Antropometria

V prípade že zo snímku nie je možné extrahovať tvár, ako je možné sledovať na obrázku 2.1, je možné danú osobu identifikovať pomocou antropometrických vlastností. Antropometria (z gréckeho slova *ánthrōpos* označujúceho človeka, a *métron* označujúceho meranie) je veda ktorá sa zaoberá meraním vlastností a schopností ľudského tela. Definuje fyzické miery veľkosti, formy a funkčných schopností človeka [11].

Antropometria sa praktizuje už od staroveku z počiatku na umelecké neskôr už aj na medicínske a lekárske účely. Veda sa začala rozširovať 17-18 storočí kedy sa vedci začali zaoberať rozdielmi medzi populáciou. V 18. storočí sa začalo objavovať čoraz viac nástrojov na merania, čo spôsobilo nárast zájmu o antropometriu. Vedci postupom času svoje postupy a spôsoby merania zlepšovali, pričom dnes sa považuje antropometria za dôležitú vedu ku ktorej je prihladané vo viacerých okruhov stavebníctva, dizajnu alebo v zdravotníctve kde dokáže spájať rôzne odvetvia medicíny, matematiky alebo štatistiky [31].

2.3.1 Prehľad antropometrických vlastností

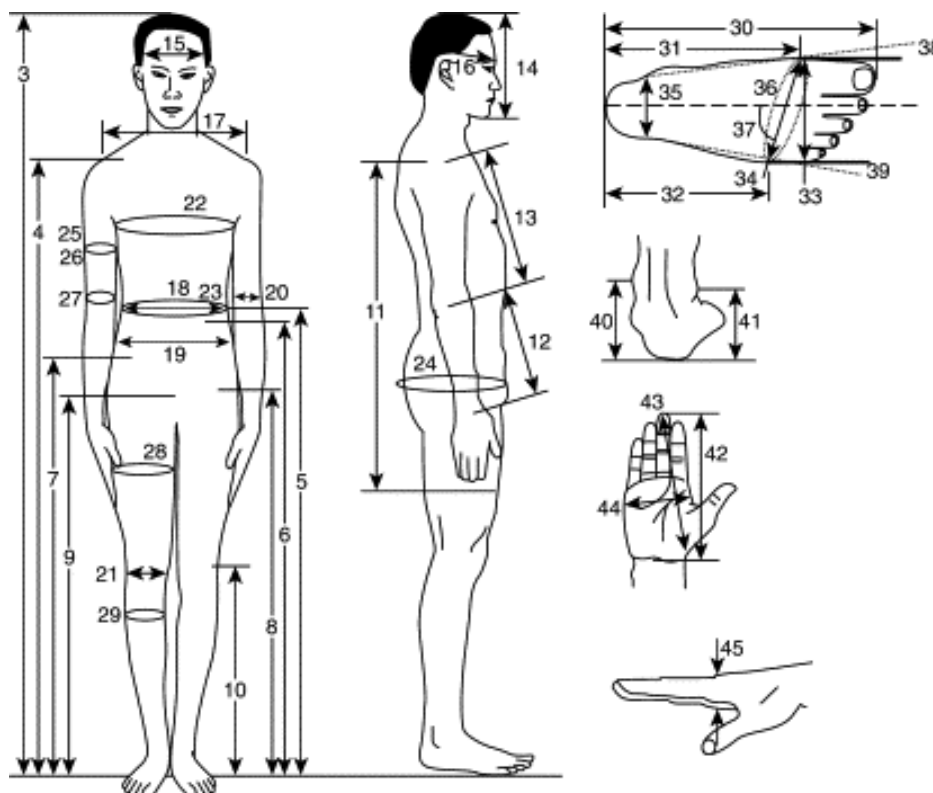
Antropometrické vlastnosti zahŕňajú kľúčové vlastnosti ľudského tela a tým aj zohrávajú kľúčovú rolu pri pochopení základných rozdielov medzi ľuďmi. Zahrňujú dôležité vlastnosti ako výška, váha, obvod hlavy, index telesnej hmotnosti (taktiež známy ako BMI (*Body mass index*)) ktoré sú taktiež popísané na obrázku 2.5. Dané antropometrické vlastnosti ktoré osoba disponuje sa vyvíjajú od počiatku života až po smrť. Veľký vplyv na antropometrické vlastnosti ma taktiež životný štýl a taktiež aj zdravotný stav 2.3.2 [11].

Kategorizovanie antropometrických vlastností a ich meraní nie je presne stanovené a rôzne zdroje si ich zoskupujú podľa svojho pohľadu alebo podľa určitej etnickej skupiny ktorej sa práve venujú. Taktiež nie je možné stanoviť najdôležitejšie antropometrické vlastnosti z toho dôvodu, že každý výskum ktorý sa venuje určitej téme si charakterizuje najdô-

ležitejšie antropometrické vlastnosti sám [18, 28].

Ako príklad antropometrických vlastností ktoré je možné relatívne jednoducho sledovať na človeku a na obrázku 2.5 môžeme zaradiť:

- Veľkosť osoby - číslo 3 na obrázku 2.5.
- Dĺžka ramena - číslo 13 na obrázku 2.5.
- Dĺžka lakťa - číslo 12 na obrázku 2.5.
- Dĺžka chodidla - číslo 30 na obrázku 2.5.



Obr. 2.5: Prehľad základných antropometrických vlastností človeka. *Prevzaté z [16].*

Antropometrické vlastnosti v oblasti tváre

Antropometrické vlastnosti tváre sa skladajú z viacerých vlastností ktoré tvár obsahuje. Ide napríklad o vzdialenosť medzi očami, výška čela, veľkosť nosu alebo veľkosť obočia. Tieto charakteristiky vedia expertovi poskytnúť mnohé informácie ohľadom etnického pôvodu osoby.

Hlavný referenčným bodom, ktorý sa využíva pri hľadaní ostatných vlastností je stred medzi očami. Z tohto bodu sa následne derivujú ostatné vlastnosti tváre ako [29]:

- **Vzdialenosť medzi stredom oboch obočí**

- **Vzdialenosť medzi stredom oka a stredom obočia**
- **Vzdialenosť medzi špičkou nosa a referenčným bodom**
- **Vzdialenosť medzi stredom úst a referenčným bodom**

Existujú ďalšie antropometrické vlastnosti ktoré ide vyčítať ľudskej tváre ako napríklad obvod hlavy, veľkosť uší a podobne. Tieto vlastnosti sú ale pre využitie na identifikovanie alebo rozpoznávanie menej užitočné a to hlavne z dôvodu že väčšina riešení sa zaoberá prednou časťou tváre z toho dôvodu, pretože sa tam nachádza najviac informačne cenných antropometrických vlastností pomocou ktorých je možné následné rozpoznávanie [29].

Pohybové charakteristiky na základe antropometrických vlastností

Pohybové antropometrické vlastnosti môžeme charakterizovať ako základné behaviorálne vlastnosti osoby. Ide o vlastnosti ktoré priamo súvisia s antropometrickými vlastnosťami ale ich vývin a prezentovanie môže daná osoba ovplyvniť.

Medzi základné pohybové vlastnosti ktoré môžeme pri ľuďoch pozorovať je chôdza. Z chôdze môžeme vyjadriť viacero ďalších vlastností ako dĺžka kroku, pohyby ramenami, rôzne deformácie kroku (hlavne pri vrodených chorobách), natočenie panvy a podobne. Medzi ďalšie pohybové vlastnosti ktoré môžeme pozorovať môžeme napríklad zaradiť spôsob pohybu s rukami poprípade náklonu hlavy.

Problémy pri vytváraní profilu na základe behaviorálnych antropometrických vlastností spočívajú v tom že osoba sa nie vždy môže pohybovať rovnako. Napríklad pri charakterizácii chôdze je náročne rozlíšiť aký veľký krok daná osoba urobí pretože sa to často odvíja práve od psychického postavenia osoby. Taktiež problémom môžu byť dočasné poranenia ktoré menia behaviorálne vlastnosti ako napríklad rôzne zlomeniny [7].

2.3.2 Vývin antropometrických vlastností

Antropometrické vlastnosti sa vplyvom času a rôznych faktorov môžu meniť. Medzi hlavný faktor zmeny antropometrických vlastností patrí starnutie osoby. Ako osoba prirodzene starne, jej antropometrické vlastnosti sa menia úmerne s vekom. Veľkú rolu pri vývine antropometrických vlastností hrá genetika, životný štýl, strava u jedinca alebo zdravotný stav jedinca. Tieto charakteristiky nám vedia poskytnúť informácie, ako rýchlo sa budú antropometrické vlastnosti u jedinca meniť.

Vývin u dospievajúcich jedincoch

Dospievanie predstavuje najvýraznejší proces zmeny antropometrických vlastností. Pri novorodencoch je možné pozorovať rýchly rast hmotnosti, výšky alebo obvodu hlavy. V období do jedného roku sa u jedinca vývoj typicky spomalí ale stále je relatívne rýchly. Do troch rokov rast ďalej pokračuje kde každoročne získajú jedinca na hmotnosti 3-4 kilogramy. Taktiež sa v tomto období mení proporčné zloženie tela čo vedie napríklad ku zmene pomeru hlavy ku zvyšku tela a končatiny sa značne predlžujú. Do 12 rokov pokračuje rast konštantným tempom.

Značné zmeny antropometrických vlastností je možné pozorovať u dospievajúcich detí v období puberty. Tá začína dievčatám typicky medzi 8. až 13. rokom života, zatiaľ čo chlapcom typicky neskôr, teda 10. až 15. rokom života, čo je možné sledovať v tabuľke 2.1.

V tomto období je typický rýchly nárast výšky, hlavne u chlapcov. Taktiež výrazne narastá svalová hmota, hmotnosť alebo svalový tuk. Tento proces trvá najčastejšie do 16. až 18. roku kedy sa vývin zastaví a z dospievajúceho jedinca sa stane dospelý jedinec [35, 4].

Tabuľka 2.1: Priemerná výška a váha pre dospievajúce deti. *Prevzaté a upravené z [10].*

Vek	Výška - Ženy (cm)	Výška - Muži (cm)	Váha - Ženy (kg)	Váha - Muži (kg)
1	68,58 - 78,74	71,12 - 81,28	6,80 - 9,07	7,71 - 9,53
2	80,01 - 91,44	81,28 - 93,98	9,98 - 14,51	10,89 - 15,42
3	87,63 - 101,60	90,17 - 102,87	11,79 - 17,24	11,79 - 17,24
4	93,98 - 107,95	95,25 - 109,22	12,70 - 19,96	13,61 - 19,96
6	106,68 - 124,46	106,68 - 124,46	16,33 - 27,22	16,33 - 27,22
8	119,38 - 137,16	119,38 - 137,16	19,96 - 36,29	20,87 - 35,38
10	127,00 - 149,86	128,27 - 149,86	24,49 - 48,08	24,49 - 46,27
12	139,70 - 162,56	137,16 - 161,29	30,84 - 61,69	29,94 - 58,97
14	149,86 - 171,45	149,86 - 176,53	38,10 - 72,57	38,10 - 72,57
16	152,40 - 172,72	160,02 - 185,42	42,64 - 78,02	47,17 - 84,37
18	152,40 - 173,99	165,10 - 187,96	45,36 - 80,74	52,62 - 91,63

2.3.3 Vývin u dospelých jedincoch

V období 20. až 30. roku života sa dostáva ľudské telo na fyzický vrchol, čo vedie k tomu že jedinec je schopný rýchlo regenerovať a zároveň má najväčšiu kostnú hustotu. V tomto období je taktiež dôležitý životný štýl ktorý jedinec dodržiava. Typické pre danú skupinu ľudí je nárast svalovej hmoty, poprípade nárast telového tuku.

V období od 30. rokoch typicky začína klesať množstvo svalovej hmoty a narastá množstvo telesného tuku a to hlavne v oblasti brucha. To isté platí taktiež aj o kostnej hustote ktorá začína klesať.

V období od 60. roku života sa zrýchľuje úbytok kostnej hmoty ako aj svalovej hmoty. Dochádza k znižovaniu výšky a to hlavne kvôli oslabeniu kostí a medzistavcových platničiek [35, 4].

Medzi ďalší faktor ktorý vedie ku vývinu antropometrických vlastností môžeme zaradiť zmenu životného štýlu u osoby. Môže ísť napríklad o rôzne diéty alebo spôsoby cvičenia. Tieto praktiky môžu viesť ku zvýšeniu svalovej hmotnosti a tým spomaliť ich úbytok v čase starnutia.

Taktiež má veľký vplyv na vývin antropometrických vlastností aj psychický stav človeka ktorý pri zlom stave môže urýchliť rýchlosť starnutia. V neposlednom rade majú významný vplyv aj choroby ktoré výrazne postihujú postavu človeka ako choroby kostí alebo svalstva [35, 4].

2.3.4 Význam antropometrických vlastností

Antropometrické vlastnosti sa v zdravotníctve využívajú na určenie fyzickej schopnosti človeka alebo predpovedať rôzne zdravotné problémy. U detí sa podľa Amerického pediatického združenia a programu hodnotenia zdravia detí a prevencie hendikepov môžu využiť tieto merania na určenie problémov so psychickým zdravím, stravovaním alebo sociálnym zázemím dieťaťa. Následne tieto merania a indikátory ktoré z týchto meraní vychádzajú

dokážu odvrátiť väčšie zdravotné problémy ktoré by mohli nastať. Môže ísť napríklad o silnú podvýživu u dospievajúceho jedinca čo vedie k tomu že jedinec sa správne nevyvíja alebo sa vyvíja pomalšie.

U športovcov dokážu tieto merania optimalizovať výkonnosť športovca tým že pomáha nastaviť správne stravovania a tým zároveň aj predchádzať problémom so stravovaním. Antropometrické fitness programy u atlétov dokážu zlepšiť kondíciu a silu športovca.

Veľký význam antropometrických vlastností sa využíva taktiež v nábytkárskom odvetví kde je nutné zohľadniť ergonómiu. Dané vlastnosti pomáhajú návrhárom navrhnuť nábytok čo najviac prívetivé k ľudskému telu aby nespôsobil krátkodobé alebo dlhodobé zdravotné problémy (hlavne s chrbticou) [23].

Veľké využitie nachádzajú antropometrické vlastnosti pri identifikácii neznámach ľudských pozostatkoch vo forenzných vedách. Pre tieto účely sa využívajú techniky ako somatometria, cefalometria, kranio-metria alebo osteometria. Vo forenzných vedách vedia tieto vlastnosti taktiež pomôcť s určením pohlavia a veku a to hlavne pomocou štatistických rozdielov medzi antropometrickými vlastnosťami ženy a muža v určitom veku [14].

2.4 Strojové učenie a extrakcia antropometrických vlastností

Ako bolo spomenuté v sekcii 2.3.1, kombinácia antropometrických vlastností dokáže jednoznačne určiť / identifikovať človeka. Avšak celková úspešnosť závisí od počtu dostupných antropometrických vlastností zo snímku. Taktiež dôležitým faktorom je kvalita a presnosť extrahovaných vlastností, ktoré by mali zaručovať spoľahlivú identifikáciu.

Na základe identifikačných vlastností vieme využiť možnosti strojového učenia a navrhnúť metódu ktorá dokáže identifikovať osobu na základe jeho stavby tela. V praxi to znamená že vieme vytvoriť profil osoby na základe množiny sledovaných vlastností ktoré dokážeme sledovať pomocou strojového učenia.

Pre tieto účely je efektívne využiť takzvané *pose* modely ktoré sú vytvorené a nacvičené pre detekciu postavy osoby. Z tohoto výstupu následne vieme extrahovať vlastnosti ako výška alebo dĺžky jednotlivých častí ľudského tela. Medzi ne patrí napríklad dĺžka lakte, chodidla, ramien a podobne.

Kapitola 3

Extrakcia antropometrických rysov postavy

Táto kapitola má za cieľ čitateľa zoznámiť ako vlastnosti vieme zachytiť z obrazu a spracovať ich do formy vhodnej pre strojové učenie. Čitateľ sa na začiatku zoznámi so strojovým učením a spôsobmi akými spôsobmi môžeme cvičiť modely strojového učenia. Ďalej budú spomenuté neurónové siete spolu s konvolučnými neurónovými sieťami ktoré majú za cieľ vytvoriť potrebný základ pre odhad pózy na obraze. Tam sa čitateľ môže dozvedieť ako prebieha detekcia osoby a detekcia pózy na obraze a aké typy detekcie poznáme. Taktiež bude predstavený príklad 2D pose modelu *YOLO Pose* a zároveň aj spôsob jeho učenia. V posledom rade bude priblížený *3D pose estimation* ktorý má za cieľ čitateľa zoznámiť s možnosťami určovania 3D pózy z obrazu.

3.1 Strojové učenie

Strojové učenie (*Machine Learning*) označuje odvetvie umelej inteligencie, ktorá prináša možnosť učiť sa z vedomostí dát a to aj bez toho aby programátor musel tieto algoritmy programovať.

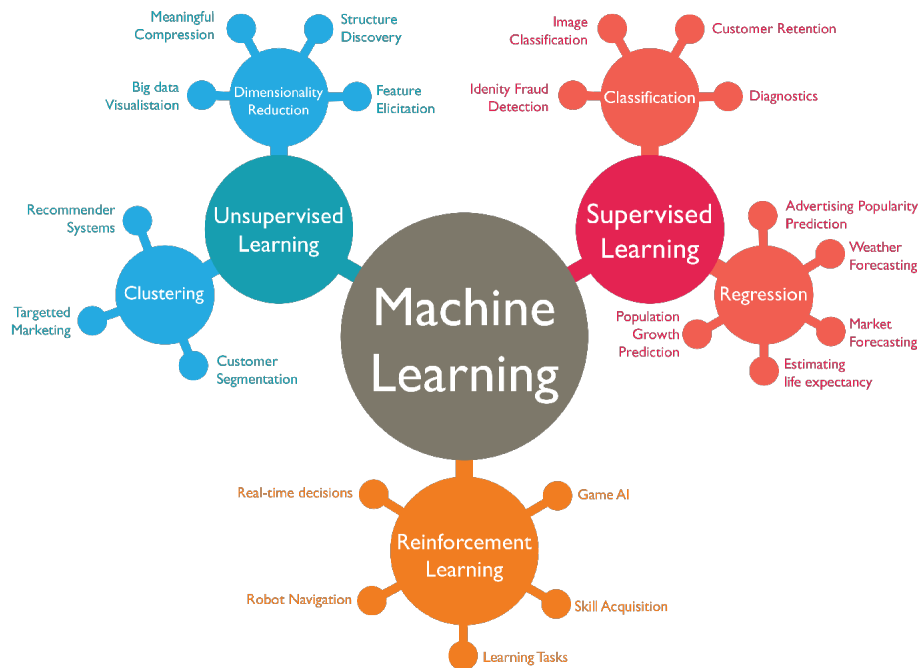
So strojovým učením súvisia 2 spolu súvisiace otázky: Ako môžeme vytvoriť systém ktorý bude mať tendenciu sa sám učiť a postupom času sám nadobúdať skúsenosti? A aké sú základné výpočtové, štatistické, teoretické zákony ktorými sa riadia systémy ako človek, spoločnosť alebo počítače?[12]. Práve tieto otázky pomáhajú programátorovi / ML expertovi navrhnuť model ktorý bude najspoľahlivejšie plniť daný účel.

3.1.1 Typy strojového učenia

Nespočet problémov ktoré strojové učenie rieši si vyžaduje rôzne druhy prístupov. Rozdiel medzi týmito prístupmi sa líši v dvoch spôsoboch: to ako dáta spracovávame a to akým spôsobom model učíme. Tieto typy učenia a zároveň aj spôsoby ich využitia je možné vidieť na obrázku 3.1.

Učenie s učiteľom

Učenie s učiteľom značí spôsob tréningu modelu na základe predom označených dát. To znamená že model sa učí s dátami pričom má dostupné odpovedajúce výstupy. Na základe



Obr. 3.1: Rozdelenie ML modelov a ich využitie. *Prevzaté z [21].*

dát sa teda vytvára model ktorý sa snaží namapovať vstupy a výstupy a tým pádom pripraviť daný model pre dáta ktoré ešte nevidel, aby pre vstupy generoval čo najsprávnejšie výstupy.

Najbežnejšie úlohy ktoré sa riešia pomocou kontrolovaného učenia sú **klasifikácia** ktorej cieľom je rozkategorizovať dáta do predom určených tried alebo štítkov alebo **regresia** ktorá zahŕňa predpoveď spojitej veličiny bez nutnosti označovania dát. Príklad klasifikácie môže byť detektor nechcených emailov alebo spamov, kde je možné model naučiť detekovať takéto správy. Naopak pri regresii predpovedáme spojitú veličinu, tú si je možné predstaviť ako predpoveď počasia, kde z prístupných dát vieme predpovedať s určitou pravdepodobnosťou ďalší vývin [26].

Učenie bez učiteľa

Učenie bez učiteľa označujeme typ učenia do ktorého nezasahuje človek. To znamená že vstup pre učenie sú neoznačené dáta. Pre učenie sa využívajú samo-učiace algoritmy ktoré sa pokúšajú nájsť vzorec medzi vstupnými dátami.

Medzi najviac používané metódy pri nekontrolovanom učení patrí **zhlukovanie** (*clustering*) ktorý rozdeľuje dáta do skupín, kde dáta majú podobné vlastnosti. Jeho hlavným cieľom je vstupné dáta rozdeliť do predom určeného počtu zhlukov a tým pripraviť model na dáta ktoré ešte nevidel.

Výhodou daného typu učenia je zbavenie sa nutnosti anotovať dáta, čo vedie k rýchlejšim a teda aj lacnejším riešeniam. Svoje uplatnenie si nekontrolované učenie nájde hlavne pri extrakcii generatívnych príznakov, identifikácii trendov alebo štruktúr [26].

Učenie so spätnou väzbou

Učenie so spätnou väzbou označuje typ algoritmu strojového učenia ktorý umožňuje agentovi učiť sa z vlastných skúsenosti z prostredia. Umožňuje vyhodnocovať optimálne správanie v konkrétnom kontexte s cieľom zlepšiť jeho efektívnosť.

Tento typ učenia je založený na princípe odmeny alebo treste, pričom jeho cieľom je využiť poznatky a skúsenosti z predchádzajúcich pokusov pri zvyšovaní odmien a minimalizácii trestov. Daný proces učenia je najčastejšie implementovaný pomocou Markovho rozhodovacieho procesu (*Markov decision process*) ktoré vyjadruje že každé rozhodnutie agenta vplýva na jeho budúce odmeny - na jeho budúci stav.

Jeho využitie je hlavne pri aplikáciach kde je potrebné rozhodovanie a adaptácia v dynamických a často komplexných prostrediach. Svoje uplatnenie nájde učenie so spätnou väzbou hlavne pri optimalizácii a automatizácii prevádzkovej efektívnosti zložitých systémov. Medzi ne patrí napríklad robotika, logistika, autonómne riadenia a podobne.[26]

3.1.2 Neurónové siete

Neurónové siete NN (*Neural Network*)[2] alebo ANN (*Artificial Neural Network*)[2] sú pokročilým nástrojom v strojovom učení ktoré využívajú koncept ľudského mozgu na riešenie zložitých problémov [2].

Základným prvkom neurónových sietí je neurón. Ten sa skladá s 4 hlavných častí [2, 20]:

- **Vstupy** - Predstavuje vstupnú bránu pre hodnoty alebo signály ktoré vychádzajú zo vstupu NN alebo iných neurónov. V tomto termíne zahrňujeme aj takzvaný **Bias** ktorý predstavuje konštantu ktorá sa započíta do súčtu vstupov.
- **Váhy** (*konceptuálne synapsie*) - Predstavujú silu spojenia medzi jednotlivými neurónmi. Negatívna váha predstavuje inhibičné spojenia¹ a naopak kladná váha predstavuje excitačné spojenia² [20].
- **Aktivačná funkcia** - Určuje či alebo ako bude neurón reagovať na sčítané váhy vstupov. Zvyčajne aktivačná funkcia je definovaná ako nelineárna funkcia ako napríklad ReLU, Sigmoid, SoftMax a podobne.
- **Výstup** - Predstavuje výslednú hodnotu aktivačnej funkcie, táto hodnota môže byť ďalej reprezentovaná ako vstup pre ďalší neurón alebo ako výstup neurónovej siete.

Z matematického hľadiska, ak $(x_1, x_2, x_3, \dots, x_n)$ sú vstupy neurónu, $(w_1, w_2, w_3, \dots, w_n)$ označujú váhy vstupov, f označuje bias a f vyjadruje aktivačnú funkciu ako Sigmoid[2], ReLu[2] a podobne[2]. Tak výstup neurónovej siete y je definovaný ako:

$$y = f(w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_n, x_n) \quad (3.1)$$

3.1.3 Konvolučné neurónové siete

Konvolučné neurónové siete (CNN) zahŕňajú prínos v tom, že redukujú počet parametrov v porovnaní s ANN (v sekcii 3.1.2). Medzi základné predpoklady u CNN patrí to, že problémy by nemali byť priestorovo závislé vlastnosťami. Práve tento predpoklad predstavuje

¹Inhibičné spojenia znižujú pravdepodobnosť že neurón vyšle signál alebo aktivuje susedné neuróny

²Excitačné spojenia zvyšujú pravdepodobnosť že neurón vyšle signál alebo aktivuje susedné neuróny

možnosti ako lepšie a efektívnejšie riešiť rôzne druhy problémov napríklad pri spracovávaní obrazu alebo hlasu.

Základným pilierom konvolučných neurónových sietí je konvulčná vrstva ktorá využíva konvolúciu. Konvolúcia spočíva v tom že sa posúva filter cez celý vstupný obraz alebo mapu prvkov. Filter má pri tom svoju veľkosť napr. 3×3 alebo 5×5 a zároveň obsahuje aj súbor váh. Pri každom posunutí filtra sa váhy filtra vynásobia s hodnotami pixelov na obraze, čím sa na výstupnej mape príznakov získa jedna hodnota.

Konvulčné vrstvy sú veľmi účinné na zachytávanie lokálnych alebo priestorových vzorov, čo ich robí účinnými prostriedkami pre aplikácie ktoré využívajú vizuálne/obrazové prostredie [2].

3.2 Odhad pózy osoby na obraze

Odhad pózy osoby na obraze môžeme popisovať ako proces kedy z digitálneho obrazu získavame orientáciu a polohu človeka alebo ľudí. Základným konceptom na ktorom odhad funguje sú kľúčové body, ktoré predstavujú časti tela na základe ktorých sa detekuje póza osoby.

Výskum ktorý prebieha ohľadom odhadu pózy je enormný, čo značia aj mnohé pokusy o využitie typov nekontrolovaného učenia (popísané v sekcii 3.1.1), ale aktuálne sú stále najlepšie výsledky z modelov ktoré sú učené čisto kontrolovaným učením (popísané v sekcii 3.1.1) [24].

Medzi základné typy modelov ktoré dokážu predikovať pózu osoby v obraze patrí odhad osoby v 2 dimenzionálnom priestore (ďalej ako 2D) a odhad pózy v 3 dimenzionálnom priestore (ďalej ako 3D). Druhé zmienené rozpoznávanie osoby v 3D priestore samozrejme nesie so sebou väčšie riziko a to hlavne z toho dôvodu, že je nutné predikovať z 2D obrazu 3D pózu z čoho plynie že model musí predikovať o jednu dimenziu navyše. Z dôvodu že pre účely rozpoznávania osoby na základe antropometrických vlastností je potrebné aby bolo možné rozpoznať dĺžky jednotlivých častí ľudského tela budeme ďalej využívať na predikcie iba modely ktoré 3D odhady umožňujú.

3.2.1 Detekcia osoby na obraze

Detekčné modely sa všeobecne začali rozširovať nástupom začiatku počítačového videnia s počiatkom 90-tych rokov. Avšak v roku 2001 bol predstavený **Viola Jones Detectors** ktorý dokázal detekovať ľudskú tvár v reálnom čase. Tento detektor sa riadil jedným z najpriamejších spôsobom detekcie, pomocou posúvných okien (*sliding windows*). Základom tohto prístupu bolo prejsť všetky miesta a mierky v obraze a sledovať či sa v danom okne nevyskytuje tvár.

V roku 2014 bola následne predstavená metóda **RCNN** (*Regions with CNN features*) ktorá rapídne zdvihla presnosť detekovania objektov na snímkoch. Metóda začína s extrakciou kandidátov na objekt, kde sa každý (kandidátny) *bounded box* preškáluje na fixnú veľkosť a vloží sa do pred-naučenej konvulčnej siete. Následne sú pomocou lineárnych klasifikátorov predikované prítomnosti objektov v snímku.

Následne bol v roku 2015 predstavený postup detekcie pomocou prvého jednostupňového riešenia **YOLO** (*You only look once*), ktorý sa od dvojestupňových detektorov spomenutých vyššie líši tým, že na celý obraz sa použije jedna neurónová sieť. Tento prístup rapídne zvýšil rýchlosť inferencie ale metóda trpí presnosťou lokalizácie hlavne u malých predmetov v porovnaní s dvojestupňovými metódami [34].

Určité modely na určovanie pózy, ktoré budú neskôr opisované využívajú pre potrebu detekovania kľúčových bodov aj oblasť objektu na ktorej bude následne prevádzať detekciu kľúčových bodov. Využívajú sa pri tom detekčné modely ako *Faster R-CNN* (*Faster Region-based Convolutional Neural Network* 3.1.3) ktorého výstup sa využíva pre ďalšie spracovávanie pose modelom.

Model si ohraničí každú osobu ktorú detekuje pomocou ohraničených polí (*bounded boxes*) v ktorých následne začína s prehľadávaním kľúčových bodov (*keypoints*). Tieto ohraničené polia sú zadané pri učení a predstavuje veľkosť a priestorovú polohu objektu v ktorej sa daný objekt (v našom prípade osoba) spolu s kľúčovými bodmi nachádza [17].

3.2.2 Detekcia kľúčových bodov

Detekcia kľúčových bodov (*keypoints*) je hlavná časť procesu odhadovania pózy osoby. Daný proces získavania bodov sa líši od architektúry siete. Medzi najznámejšie prístupy pre detekciu patrí **Metóda z dola na hor** a **Metóda zhora nadol** (vizualizácia metódy na obrázku 3.2).

Metóda z dola na hor

Proces začína vstupom obrazu do neurónovej siete pred-spracovaním. Tento proces zahŕňa normalizáciu a škálovanie ktoré zabezpečuje aby vstupný snímok mal požadovanú veľkosť pre vstup do neurónovej siete.

Následne obraz prejde cez extraktor vlastností (*feature extractor*) ktorý sa skladá z viacerých vrstiev konvolučných neurónových sietí 3.1.3. Výsledkom tejto časti je takzvaná heat mapa alebo mapa vlastností ktorá reprezentuje detekované vlastnosti na snímku. Týchto máp môže byť viacero - každá hlava neurónovej siete reprezentuje jeden keypoint alebo iba jedná mapa ktorá reprezentuje všetky sledované kľúčové body. Daná heat mapa reprezentuje pravdepodobnosť s ktorou sa na obraze kľúčový bod nachádza.

Následne sa tieto heatmapy spracujú pričom sa vyhodnotia výsledné kľúčové body ktoré majú pravdepodobnosť istoty väčšiu ako threshold^3 . Tento typ detekcie sa vyskytuje pri modeloch ako *OpenPose*, *PoseNet*, *AlphaPose* [13].

Metóda zhora nadol

Metóda zhora na dol sa od metódy pomocou tepelných máp (Metódy zdola na hor) líši spôsobom prístupu k objektu, na ktorom kľúčové body hľadá. Tento spôsob narozdiel od tepelných máp v jednom kroku detekuje ohraničenia osôb 3.2.1 a zároveň aj pózu osoby, presnejšie kľúčové body osoby. Metóda zhora nadol pre tento postup využíva **priamu regresnú detekciu**. Tento postup obnáša spájanie kľúčových bodov objektu priamo s kotvami (*anchors*) ktoré označujú preddefinované polia používané na detekciu objektov v rôznych mierkach a pomeroch strán [17].

Celkový proces detekcie je oproti postupu s tepelnými mapami zjednodušený. Postup sa vo veľkej miere spolieha na ohraničenia osôb čo vedie k tomu, že detekcia u snímokoch kde nie je dobre rozpoznateľný objekt nemusí byť presná alebo nemusí byť vôbec. Na druhú stranu, výhoda daného postupu spočíva v lepšom mapovaní kľúčových bodov na objekt a to vďaka kotvám, ktoré umožňujú že model bude lepšie môcť detekovať kľúčové body na

³Pod pojmom **threshold** (*slov. prah*) sa vyznačuje v ML konštanta ktorá predstavuje kritický bod ktorý určuje ako bude dané rozhodovanie klasifikované.

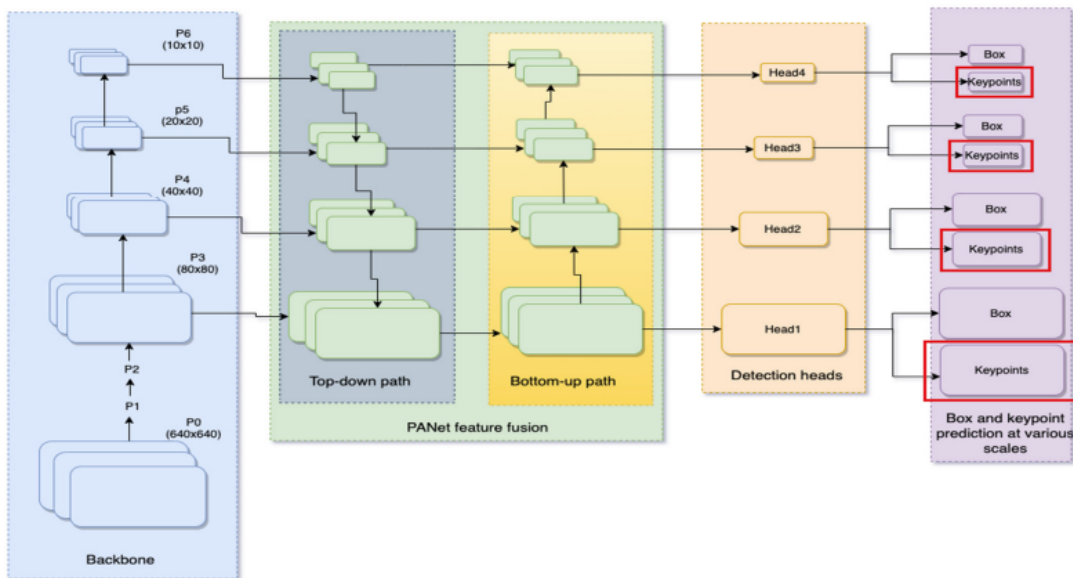
objekte.

Pre efektívnu a presnú detekciu je nutné precíznejšie tréningovanie čo sa vyznačuje kvalitnejšími dátami a väčším množstvom dát. Medzi hlavné výhody využitia metódy zhora na dol patrí vysoká efektivita u detekcie póz na scénach kde sa nachádza väčší počet osôb. Táto výhoda však vedie aj k opačnému problému a to s problémom **oklúzie** (*occlusion*). Tento výraz označuje problém detekovania kľúčových bodov ktoré sa nenachádzajú v ohraničení objektu [17].

Tento typ detekcie sa využíva hlavne pri návrhu sieťovej architektúry, ale aj ako časť u robustnejších architektúrach ako *YOLO-Pose*.

3.2.3 YOLO Pose

YOLO (*You only look once*) Pose označuje metódu v oblasti strojového učenia a počítačového videnia pre detekciu pózy v obraze. Táto derivácia modelov YOLO existuje vo viacerých variantách podľa toho aký detekčný model sa využije ako základný (*base model*) pre model detekujúci pózu. Ako základný model môžu byť použité modely ako YOLOv5, YOLOv6, YOLOv8 alebo iné detekčné modely typu YOLO.



Obr. 3.2: Yolo pose architektúra ktorá je založená na YOLOv5. *Prevzaté z [17].*

Model sa pri detekovaní zameriava na 80 tried, ktoré sú definované podľa štandardu COCO-Pose 3.2.3. Podľa týchto tried teda pripadá na jednu kotvu 85 elementov ktoré vyznačujú:

- 4 elementy pre ohraničenie objektu
- 1 element pre skóre objektu ⁴

⁴Metrika ktorá udáva istotu pri určovaní ohraničenia objektu

- 80 elementov pre označenie dôvery-istoty (*confidence*) pre každú z 80 tried.

Podľa mriežky sú následne pre každú časť priradené 3 kotvy iných rozmerov [17].

V prípade toho že chceme využiť Pose model čisto iba na detekovanie jedného objektu - osoby, využijeme jednu triedu kde budeme určovať n kľúčových bodov predstavujúcich určitú časť ľudského tela. Z tohoto pohľadu podľa štandardu COCO a anotácii v rámci COCO datasetu môžeme počítať s **17 kľúčovými bodmi**. Podľa predchádzajúcich anotácii máme 51 bodov identifikovaných následovne: $\{x, y, conf\}$, kde x a y označujú súradnice kľúčového bodu a $conf$ označuje dôveryhodnosť alebo istotu s ktorou je bod detekovaný. Z tohto hľadiska teda pre každú kotvu predikuje hlava kľúčových bodov 51 elementov a hlava boxu (ohraničenia objektu) predikuje ďalších 6 elementov. To znamená že výsledný vektor pre n kľúčových bodov je rovný [17]:

$$P_v = \{C_x, C_y, W, H, Box_{conf}, Class_{conf}, K_x^1, K_y^1, K_{conf}^1, \dots, K_x^n, K_y^n, K_{conf}^n\} \quad (3.2)$$

kde C_x a C_y vyznačujú stredové súradnice ohraničenia objektu, W a H označujú šírku a výšku ohraničenia objektu, Box_{conf} označujú dôveryhodnosť alebo istotu že objekt sa nachádza v ohraničení objektu, $Class_{conf}$ označuje dôveryhodnosť/istota že objekt bol správne klasifikovaný, K_x a K_y označujú x a y súradnice predikovaného kľúčového bodu, K_{conf} označuje dôveryhodnosť/istota že kľúčový bod bol správne identifikovaný a $\{K_x^m, K_y^m, K_{conf}^m\}$ označuje trojicu hodnôt identifikujúcich kľúčový bod m [17].

Dôveryhodnosť kľúčového bodu sa pri tréovaní určuje podľa vlajky kľúčového bodu - viditeľnosť. Tá označuje do akej miery je daný bod na scéne viditeľný. Naopak, pri inferencii⁵ modelu označuje do akej miery si je model istý že sa tam kľúčový bod skutočne nachádza. Pokiaľ sa istota nedovíši hodnotu thresholdu, kľúčový bod bude ignorovaný a nebude súčasťou predikcie na výstupe [17].

Tréovanie a štandard COCO pose

Štandard COCO Pose označuje formát vstupných anotácii ktoré sú pomocou *Data loader-u* vkladané do modelu pre proces učenia alebo inferencie. *Data loader* sa stará o prenos dát z úložiska do pamäte kde sú dáta umiestnené zároveň aj s tréovaným modelom. Nad dátami prevádza **transformáciu** - škálovanie, normalizácia, augmentácia a podobne, **grupovanie** - usporiadúvanie dát do skupín ktoré predstavujú skupinu dát pre tréovanie, **paralelné spracovanie** - príprava ďalšej skupiny dát, zatiaľ čo predchádzajúca skupina sa spracováva a ďalšie funkcie ktoré sa odvíjajú od vstupných dát a modelu ktorý je použitý [19].

Data Loader pre YOLO Pose akceptuje anotácie ktoré sú vo *Ultralytics YOLO formate* ktorý je zhodný s formátom anotácii ktorý využíva štandard COCO pose. Tento formát definuje vstupné dáta 2-mi rôznymi spôsobmi. Dáta môžu byť reprezentované dvoj dimenzionálne $\{K_x, K_y\}$ kde dáta obsahujú iba x a y súradnice bodu alebo trojdimenzionálne $\{K_x, K_y, Visibility\}$ kde sa nachádza okrem x a y súradnice aj hodnota viditeľnosti ktorá

⁵Pojem označujúci nasadenie modelu a jeho následovný beh.

určuje do akej miery je bod na scéne viditeľný [19, 8].

Anotácie musia byť uložené pod rovnakým názvom ako vstupný snímok pričom formát súboru, jedného objektu na snímku, pri počte kľúčových bodov n musí byť:

pre 2-dimenzionálne:

```
<trieda> <x-stred> <y-stred> <šírka> <výška> <x_1> <y_1> ... <x_n> <y_n>
```

alebo pre 3-dimenzionálne:

```
<trieda> <x-stred> <y-stred> <šírka> <výška> <x_1> <y_1> <viditeľnosť_1> <x_n>  
<y_n> <viditeľnosť_n>
```

Všetky hodnoty až na triedu, musia byť v normalizovanom tvare, t.j. od 0 po 1. Dáta pri tréningu musia obsahovať aj *YAML* súbor ktorý obsahuje všetky informácie potrebné pre data loader pre správne načítanie snímok a anotácii [8].

3.3 Detekcia pózy osoby v 3D

V minulej kapitole bola popísaná estimácia pózy v 2D priestore. Avšak pre potreby rozpoznávania osôb na základe antropometrických vlastností je nutné aby bolo možné detekovať pózu v troch dimenziách. Aktuálne je toto odvetvie veľmi populárne vo výskumoch kde nachádza svoje uplatnenie či už v zdravotníctve, analýze športov alebo pri tvorbe filmov (*motion capture*).

Každopádne estimácia pózy v 3D prináša do výskumov aj veľké výzvy a to z toho dôvodu že pri procese estimácie je nutné z dvoj-rozmerného obrazu získať troj-rozmernú postavu. To znamená že zatiaľ čo 2D estimácia rieši problémy ako variácie polohy tela, komplikované pozadia, rôzne farby a dizajny oblečenia 3D estimácia rieši ďalej problémy ako nízky počet dát z reálneho sveta, nejednoznačnosť hĺbky obrazu alebo obrovské nároky na polohu. Tieto a ďalšie problémy budú rozpisované v kapitole 3.3.1 [32].

3.3.1 Výzvy pre detekciu pózy v 3D

Ako bolo spomenuté v 3.3, detekcia pózy v 3D si odnáša väčšinu problémov ktoré musí riešiť aj detekcia v 2D a zároveň sa pridávajú aj ďalšie výzvy ktoré musí detekcia v 3D riešiť. Medzi základné tri problémy ktoré musí detekcia pózy v 3D riešiť patria [32]:

- **Rôzne vstupy** - V praxi to znamená, že kvôli nejednoznačnosti vstupu je komplikované určiť presnú pózu. V reálnom svete sa na vstupe môžu zobrazovať rôzne vizuálne podnety ktoré môžu komplikovať správne určenie. V tom prípade pri detekcii 3D pózy na základe 2D pózy môže viesť ku veľkým problémom a výstupy sa môžu stať nejednoznačné a nepresné.

Ako riešenie by mohlo byť časový zber dát, kde je možné rozlíšiť rôzne vizuálne podnety a určiť ten správny na detekciu pózy [32].

- **Problém viac osôb** - Oddelovanie pózy viacerých osôb je jeden z problémov ktoré musí riešiť aj detekcia v 2D. Pri detekcii v 3D sa ale tento problém stáva viac problematický a to hlavne pri snímkoch kde sú osoby veľmi blízko seba. Tento problém je riešiteľný pomocou snímania obrazu z viacerých uhlov a tým predísť snímkam kde sa osoby prelínajú [32].

- **Problém s dátami v reálnom svete** - Pre dáta na 2D estimáciu je možné využiť dáta z reálneho sveta a ďalej ich iba označiť na tréning. Pri 3D estimácii však nastáva problém že tento proces je oveľa viac zložitý a dátové sady, ktoré sú využité pre tréning *pose modelov*, využívajú často dáta získané v dokonalých študijných podmienkach. To vedie k tomu že dáta a zároveň aj model sú veľmi zovšeobecnené čo vedie k horším výsledkom v podmienkach reálneho sveta [32].

3.3.2 Modelovanie pózy

Z dôvodu že ľudské telo je veľmi komplexné, existujú taktiež aj viaceré možnosti ako môžeme pózu osoby modelovať. Tento výstup sa líši podľa toho aký model využívame a ako chceme ďalej výstupy spracovávať.

Model založený na kostre

Tento typ modelu je veľmi používaný pri estimácii pózy v 2D priestore. Z toho dôvodu bol taktiež aj rozšírený do 3D. Ľudské telo a teda aj jeho póza je opísaná stromovou štruktúrou ktorá obsahuje určitý počet bodov, ktoré predstavujú rôzne časti tela [32].

SMPL model

SMPL model (*Skinned Multi-Person Linear model*) sa snaží vyjadriť pózu pomocou zobrazenia celého tela. Ľudská koža je zobrazená ako sieť tvorená trojuholníkmi, ktorá má 6890 bodov (vrcholov). Táto sieť je nastaviteľná podľa parametrov, ktoré určujú tvar a pózu (postoj) tela. Parametre tvaru sú použité na vytvorenie modelu, ktorý zodpovedá rôznym telesným proporciám, ako sú výška a hmotnosť, zatiaľ čo parametre pózy slúžia na vytvorenie modelu špecifických deformácií tela. Pozície trojrozmerných póz (polôh) tela možno určiť vypočítaním týchto parametrov tvaru a pózy tela. Príklad výstupu SMPL modelu je možné pozorovať na obrázku 3.3 [32].



Obr. 3.3: Príklad SMPL modelu vyjadrujúceho ľudskú pózu. Biele body na obrázku vyjadrujú pred definované body. *Prevzaté z [32].*

3.3.3 Prístupy k detekcii pózy

V tejto sekcii sa čitateľ môže dozvedieť aké spôsoby a metódy sa využívajú pre detekciu pózy v 3D. Existuje viacero spôsobov prístupu k detekcii póz podľa toho či využívame viacero snímok z viacerých uhlov alebo využívame iba snímky z jedného pohľadu. Taktiež vieme rozdeliť metódy podľa prístupu detekcie - priama detekcia alebo získavanie 3D pózy z 2D.

Priama estimácia

Priama detekcia pózy funguje na princípe konečných neurónových sietí, ktoré priamo mapujú 3D body na vstupný obraz. Túto metódu môžeme ďalej rozdeliť na pod metódy podľa toho akým spôsobom toto mapovanie prevádzajú [32, 24]:

- **Detekčné metódy** - Táto metóda pracuje na princípe predpovedania tepelných máp (*heatmap*) pre každý kĺb. Táto tepelná mapa predstavuje s akou pravdepodobnosťou sa daný bod nachádza na snímku v 3 rozmernom priestore. Jedna z možných implementácií je Pavlakosová metóda ktorá využíva objem (*volume*) na reprezentáciu 3D pózy a následne spočíva v trénovaní konvolučnej neurónovej siete (popísané v sekcii 3.1.3) na predpovedanie voxelovej (*volumetric pixel*) pravdepodobnosti pre každý bod v objeme. Tieto metódy využívajú postupné zvyšovanie rozlíšenia heatmapy a iteratívne moduly odhadu pre zlepšenie presnosti [32, 24].
- **Regresné metódy** - Ich cieľom je priamo odhadnúť polohy bodov bez vytvárania tepelných máp keďže aj odhad polohy človeka je v podstate regresný problém. Využíva pri tom neurónovej siete, ktorá sa učí prevádzkať vstupné obrazy na súradnice kĺbov v 3D priestore. Príkladom sú práce od *Li a Chan (2014)* a *Zhou et al. (2016)*, kde sa používajú rôzne architektúry sietí na priame odvodenie 3D pózy z obrazov [32, 24].

Zdvíhanie z 2D na 3D

Táto metóda sa zameriava na využitie výsledkov z estimácie pózy v 2D pričom sa 2D údaje transformujú do 3D. Tieto pokusy vznikli z dôvodu toho aby bolo možné zlepšiť presnosť estimácie v reálnom svete. Proces zahŕňa zdvíhanie alebo zjednocovanie tepelných máp z 2D estimátorov a 3D obrazovými podnetmi.

Tieto metódy majú obrovský potenciál v reálnom svete a to z dôvodu že nepotrebujú priame 3D dáta na učenie [32].

3.3.4 Metriky pre výpočet chyby

Pre správne určenie bodov je potrebné poznať s akou presnosťou daný model vie daný model predikovať výstup. Pre tento účel sa využíva viacero metrík pričom dvoje hlavné sú rozpísané nižšie:

Mean Per Joint Position Error

Táto metrika počíta priemernú chybu polohy pre každý kĺb. MPJPE sa získava ako priemerná vzdialenosť medzi odhadnutou a skutočnou (*ground truth*) pozíciou každého bodu. Tieto body sú zohľadnené a výsledné MPJPE hodnoty sú priemerované cez všetky snímky. Všeobecne platí, že čím je hodnota chyby *menšia*, tak model dokáže presnejšie predikovať výslednú pózu. Vzorec pre výpočet je nasledovný [32]:

$$\text{MPJPE} = \frac{1}{N} \sum_{i=1}^N \left\| P_{\text{est}}^{(i)} - P_{\text{gt}}^{(i)} \right\|_2 \quad (3.3)$$

kde N je počet predikovaných bodov, $P_{\text{est}}^{(i)}$ je odhadnutá poloha bodu, $P_{\text{gt}}^{(i)}$ je skutočná poloha bodu a $\|\cdot\|_2$ je euklidovská norma - vzdialenosť medzi dvoma bodmi v priestore.

Percentage of Correct Keypoints

Metrika meria percento správne určených bodov. Bod sa pri tom považuje za správne určený pokiaľ jeho vzdialenosť v priestore je menšia ako určitý prah. Všeobecne taktiež platí, že čím je percento správne určených bodov *vyššie*, model dokáže lepšie predikovať výstupnú pózu. Výpočet spočíva vo vzorci [32]:

$$\text{PCK} = \frac{\text{Počet správne určených bodov}}{\text{Celkový počet bodov}} \quad (3.4)$$

Kapitola 4

Príprava dát a rozpoznávanie osôb s využitím strojového učenia

Táto kapitola sa zameriava na dve kľúčové časti : extrakciu dát a vytvorenie ML modelu ktorého cieľom je správne klasifikovať osoby na základe ich antropometrických vlastností. Pri extrakcii dát je taktiež potrebné aby dáta boli čo najviac validné a aby zo snímku bolo extrahované čo najviac charakteristík pre potreby učenia a testovania modelu. Cieľom tejto kapitoly je vytvoriť model pomocou strojového učenia, ktorý tieto charakteristiky extrahované zo snímku čo najlepšie "pochozí" a dokáže čo najviac efektívne a čo najviac správne predikovať identitu osoby.

4.1 Návrh riešenia

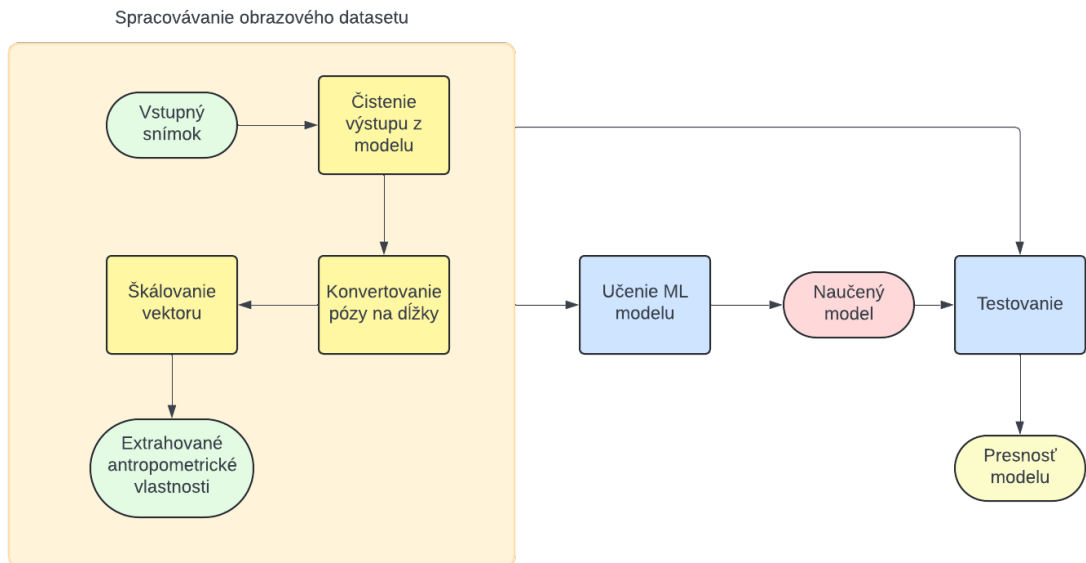
Prvá etapa návrhu spočíva v extrakcii dát, kde sú obrazové dáta konvertované pomocou *pose modelu* na množinu bodov v priestore, ktoré charakterizujú pózu osoby. Táto póza je následne konvertovaná na množinu dĺžok. Tieto dĺžky reprezentuje tabuľka dát ktorá nesie antropometrické vlastnosti charakterizujúcu osobu. Po extrakcii sú tieto dáta využité pre tréning modelu. Návrh riešenia je možné vidieť na obrázku 4.1.

Pri testovaní sa bude postupovať rovnakým spôsobom extrakcie charakteristík ako pri učení s tým rozdielom, že dáta nebudú obsahovať informáciu o osobe, ktorej dané antropometrické vlastnosti patria. Na základe počtu správne identifikovaných osôb a celkového počtu následne dokážeme určiť presnosť modelu, ktorú budeme považovať ako mieru presnosti modelu.

4.2 Dátové sady

Pre potreby vytvorenia systému ktorý je schopný detekovať osoby na základe antropometrických proporcií je nutné aby obrazová dátová sada splňovala základné kritéria:

- Snímky musia byť vo vysokom rozlíšení.
- Na snímku sa nachádza iba jedna osoba.
- Osoby sú jasne identifikovateľné/rozlíšiteľné.
- Prevažujú snímky celého tela nad odstrižkami u jednej osoby.



Obr. 4.1: Návrh riešenia s využitím strojového učenia.

Pre tieto účely boli vybrané základné 2 verejne dostupné dátové sady ktoré najlepšie spĺňajú tieto podmienky. **LaST Dataset** [27] a **Player Reidentification Challenge Dataset** [33].

4.2.1 Large-Scale Spatio-Temporal Person Re-identification Dataset (*LaST*)

LaST dataset je mohutný reidentifikačný dataset ktorý pozostáva z veľkého množstva dát, 10 862 oddelených osôb. Tieto osoby sú prevažne vystrihnuté z filmov čo spôsobuje že snímky majú nízku kvalitu a prevažné množstvo snímok osôb nemajú na snímkoch plne viditeľné celé telo. Počet snímok je priemerne 14,59 unikátnych snímok na jednu osobu [27].

4.2.2 Player Reidentification Challenge Dataset

Player Reidentification Challenge Dataset na rozdiel od LaST datasetu pozostáva z menšieho množstva unikátnych osôb 486. Osoby na snímkach sú športovci zachytený počas zápasov rôznych športov čo poskytuje lepšiu diverzitu dát - osoby sú na snímkoch v rôznych polohách. Taktiež je v tejto dátovej sade lepší pomer medzi osobami s celým telom a osobami s odstrihnutým telom. Snímky sú priemerne ešte v nižšej kvalite ako u predchádzajúceho datasetu čo môže zhoršiť presnosť výsledkov z modelu pre estimáciu 3D pózy.

Oproti predchádzajúcemu datasetu, dataset obsahuje priemerne vyšší počet unikátnych snímok na jednu osobu 19,65 [33].

4.3 Spracovávanie obrazového datasetu

Dáta získané z verejných obrazových dátových sád obsahujú dôležitú časť procesu ale pre získanie potrebných antropometrických vlastností je nutné aby obrazová dátová sada bola konvertovaná do tabulkovej dátovej sady ktorá bude dané antropometrické vlastnosti osoby reprezentovať.

Pre tento účel bol práve použitý 3D estimátor pózy na obraze (popísaný v sekcii 3.3) ktorý dokáže pózu v obraze predpovedať a na základe týchto dát danú dátovú sadu vytvoríť.

4.3.1 Model pre estimáciu 3D pózy

Pre účely získavania antropometrických vlastností je potrebné využiť model ktorý predikuje pózu v 3 dimenziách 3.3. Taktiež je potrebné aby model predikoval kostru na základe ktorej dané antropometrické vlastnosti vieme extrahovať 3.3.2. Z tohto hľadiska je aktuálne možné využiť len *Open Source* model MeTRAbs [25].

MeTRAbs Absolute 3D Human Pose Estimator

MeTRAbs model je ktorý využíva priamu estimáciu 3.3.3 pre detekciu pózy v priestore. Model dokáže v snímku detekovať osoby, ich pózu v 2D priestore a pózu v 3D priestore [25].

Výhodou použitia tohto modelu je zabudované pre-procesovanie dát kedy si knižnica modelu sama snímok upraví pre potreby modelu. Taktiež výhodou je široký výber kostier ako SMPL, H36 alebo COCO. Model má taktiež zabudovanú supresiu póz ktoré sú pre osobu nepravdepodobné, čo zabezpečuje lepšiu konzistenciu výstupných dát [25].

Skeleton

Ako skeleton ktorý bude kostra generovať bol zvolený *simpl+head_30* ktorý obsahuje 30 kľúčových bodov. Tieto kľúčové body pokrývajú celú oblasť tela od nôh, trupu, rúk až po hlavu. Príklad pózy vyjadrenej pomocou skeletonu SMPL je možné vidieť na obrázku 4.2.

Body ktoré kostra popisuje majú len približný charakter. To znamená, že body nepopisujú konkrétnu kosť ale iba približné miesto kde sa daná časť tela nachádza. Konkrétne body ktoré kostra *simpl+head_30* popisuje je možné vidieť v tabuľke 4.1.

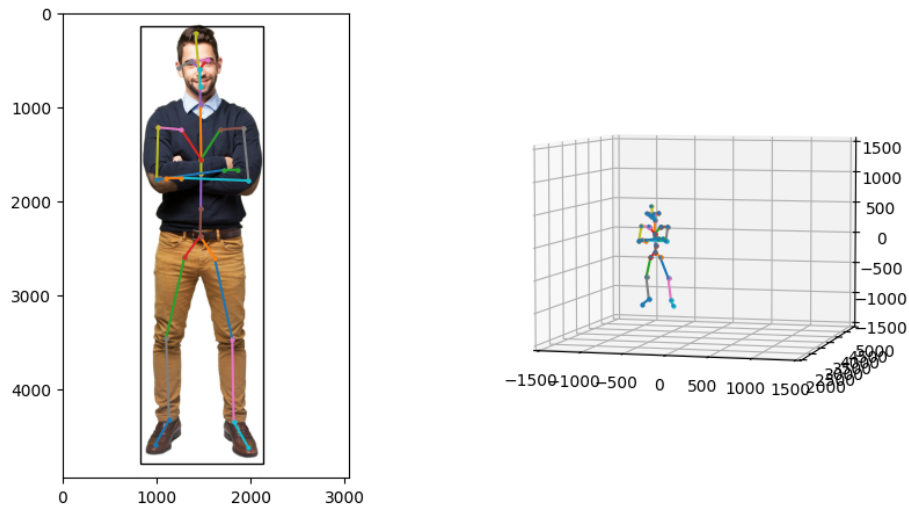
4.3.2 Spracovanie výsledkov

Póza získaná z datasetu nesie informácie o polohe častí tela v priestore. Tieto dáta však na účely strojového učenia nie sú vhodné a z toho dôvodu je ich nutné konvertovať do priaznivejšieho formátu.

Ako nosnú informáciu ktorá by mala identifikovať osobu sa bude označovať vektor vzdialeností medzi jednotlivými kľúčovými bodmi. Ide napríklad o vzdialenosť zápästia po lakeť alebo dĺžka stehennej kosti (popísané v sekcii 2.3.1). Tieto vlastnosti pokrýva práve skeleton, ktorý body medzi ktorými sa bude vzdialenosť merať definuje.

Čistenie výstupu z modelu

Model produkuje pózu celého tela aj pokiaľ sa celá časť tela na obraze nevyskytuje. Z toho dôvodu je potrebné dané body nájsť a odstrániť ich pre potrebu čo najväčšieho obmedzo-



Obr. 4.2: Príklad inferencie modelu MeTRAbs. Vľavo osoba a jej póza v 2D. Vpravo póza osoby v 3D priestore.

Tabuľka 4.1: Body kostry *simpl+head_30* a ich význam na ľudskom tele.

Bod	Význam	Bod	Význam	Bod	Význam
pelv	panva	ltoe	ľavý prst na nohe	lwri	ľavý zápästie
lhip	ľavý bok	rtoe	pravý prst na nohe	rwri	pravý zápästie
rhip	pravý bok	neck	krk	lhan	ľavá ruka
bell	brucho	lcla	ľavá kľúčna kosť	rhan	pravá ruka
lkne	ľavé koleno	rela	pravá kľúčna kosť	nose	nos
rkne	pravé koleno	head	hlava	leye	ľavé oko
spin	chrbtica	lsho	ľavé rameno	lear	ľavé ucho
lank	ľavý členok	rsho	pravé rameno	reye	pravé oko
rank	pravý členok	lelb	ľavý lakeť	rear	pravé ucho
thor	hrudník	relb	pravý lakeť	htop	vrchol hlavy

vania šumu v dátach. Táto nežiadúca situácia sa dá odstrániť pomocou algoritmu ktorý porovná body pózy v 2D priestore či sa nachádzajú vo vnútri boxu ktorý ohraničuje danú osobu¹ a do výsledku propaguje iba tie 3D body, ktorého 2D bod sa nachádza vo vnútri boxu nachádza.

Konvertovanie pózy na dĺžky

Body ktoré sa nachádzajú v obraze je následne nutné previesť do podoby vektora dĺžok. Tento vektor obsahuje vzdialenosti medzi dvoma bodmi ktoré sú definované podľa zvolenej kostry. Pre výpočet vzdialenosti v kartézskom priestore je využitá euklidovská vzdialenosť

¹Vychádzame z predpokladu že model predikuje box okolo osoby iba v obraze

ktorá je definovaná nasledovne:

$$d = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (4.1)$$

kde N je veľkosť dimenzie (v danom prípade 3), x je prvý bod (v danom prípade ako x , y , z) a y je druhý bod (v danom prípade ako x , y , z).

Tabuľka 4.2: Ukážka extrakcie antropometrických vlastností z procesu pred-spracovania dát z dátovej sady LAsT. Mená stĺpcov vyznačujú vzdialenosti medzi jednotlivými časťami popísaných v tabuľke 4.1 oddelených podtržníkom.

Person ID	Lhip_lkne	Lhip_pelv	Rhip_rkne	Rhip_pelv	...	Head_nose
000009	345.86	96.21	-1.00	94.04	...	104.09
000009	-1.00	87.29	287.32	95.56	...	89.81
000009	-1.00	92.69	-1.00	83.00	...	90.05
000009	-1.00	92.07	-1.00	100.59	...	97.19
000016	-1.00	97.25	-1.00	97.81	...	83.32

Škálovanie vektoru

Body pózy ktoré sú vyprodukované modelom sa v 3D priestore môžu nachádzať kdekoľvek. Závisí to podľa toho ako je vzdialená kamera od postavy alebo v akých podmienkach bola snímka zachytená. Z toho dôvodu je nutné si definovať dĺžku ktorá bude slúžiť ako referenčná vzdialenosť dvoch bodov.

Model ktorý je využitý pri estimácii kostry využíva základný koreňový bod (*pelv*) od ktorého sa následne generujú ďalšie body. Z toho vyplíva že práve tento bod bude najideálnejším kandidátom na prvý bod pre škálovanie. Ako ďalší bod ktorý môže byť využitý pre referenčnú vzdialenosť je pupok. Nachádza sa priamo nad koreňovým uzlom čo vedie k tomu, že jeho presnosť by mala byť vyššia a tým pomocou škálovania neznehodnotiť ostatné informácie z pózy.

4.3.3 Analýza výsledných dát

Pri analýze dát je cieľom zistiť do akej miery sú dáta vhodné na strojové učenie. Pre tieto potreby budú použité viaceré metriky, u ktorých budeme analyzovať dáta extrahované pomocou *3D pose estimator*.

Na tabuľke výsledných dát 4.2 je možné si všimnúť veľký výskyt nezistených dát (-1). Tieto údaje sa buď modelu nepodarilo predikovať alebo sa časť pózy na snímku nevyskytovala. Dôležitým parametrom analýzy bude teda zistiť, akému percentu dát nechýbajú žiadne informácie alebo mu chýbajú 5 a menej informácií.

Pri analýze sa bude ďalej skúmať **triedna(vnútorňú) variabilita dát**, ktorú v daných tabuľkových dátach budeme označovať ako rozsah, v akom sa údaje u jednej triedy líšia od seba navzájom. Presnejšie to bude značiť ako sa jednotlivé dĺžky líšia od seba pri jednej

osobe. Pomocou tejto metriky vieme určiť aká veľká je odchýlka dát v jednej triede (osoby) - čím menšia odchýlka, tým lepšie.

Medzitriedna(vonkajšia) variabilita dát bude naopak určovať rozsah, v akom sa údaje jednotlivých výstupných bodov líšia od seba. Použitím tejto metriky vieme určiť aká je veľká odchýlka medzi jednotlivými triedami - čím väčšia odchýlka tým lepšie.

Analýza dát z datasetu LAsT

Pri meraní bolo použitých prvých 25 identít. Tieto identity boli následne kontrolované na počet chýbajúcich hodnôt. Z 8874 celkových hodnôt bolo chýbajúcich 1626 čo predstavuje 18,32% z celkového počtu hodnôt. Zároveň z 306 záznamov, kde jeden záznam značí jeden snímok, je 5 záznamov ktoré obsahujú všetky hodnoty čo je 1,63% z celkového počtu. Záznamov ktorým chýba 5 alebo menej hodnôt, je následne 142 čo predstavuje 46,41% z celkového počtu 306 záznamov.

Triedna(vnútoraná) variabilita u datasetu bola získaná spôsobom výpočtu štandardnej odchylky pre každý atribút. Odchýlky atribútov sa spriemerovali a výsledok sa zaokrúhlil na 4 desatinné miesta, čo viedlo ku hodnote **8,9877**.

Pre medzitriednu(vonkajšiu) variabilitu bol najskôr pre každú identitu vypočítaný aritmetický priemer každého atribútu. Následne sa medzi týmito hodnotami vypočítala štandardná odchýlka. Tieto hodnoty boli použité pre výpočet aritmetického priemeru a výsledok bol zaokrúhlený na 4 desatinné miesta, to viedlo ku hodnote **7,2489**.

Analýza dát z datasetu Player Reidentification Challenge

Rovnako ako aj pri analýze datasetu LAsT v sekcii 4.3.3, pre merania bolo použitých 25 unikátnych osôb. Pri celkovom počte 13340 celkových hodnôt, bolo chýbajúcich 1073 čo predstavuje 8,04% zo všetkých hodnôt. Celkový počet záznamov 25 osôb bol 460 kde 49 záznamov obsahovalo všetky hodnoty, čo predstavuje 10,65%. Pokiaľ ide o záznamy kde počet chýbajúcich hodnôt je rovný alebo menší ako 5, tak počet takých záznamov je 413 čo predstavuje 89,78% zo všetkých záznamov.

Triedna variabilita u datasetu bola meraným rovnakým spôsobom ako pri datasete LAsT v sekcii 4.3.3. Pri zaokrúhlení na 4 desatinné miesta vychádza vnútorná variabilita **7.8712**.

Podobne ako pri triednej variabilite, aj medzitriedna variabilita bola meraná rovnakým spôsobom. Po zaokrúhlení výsledku vychádza medzitriedna variabilita na hodnotu **7.2489**.

Vyhodnotenie analýzy

Analýza dát pomocou týchto dvoch metrik ukázala že triedna variabilita dát je väčšia ako medzitriedna variabilita dát. Túto situáciu môžeme interpretovať tým spôsobom, že dáta v jednej triede sú od seba vzdialené viac ako dáta medzi jednotlivými triedami. To spôsobuje že s danými dátami je nižšia pravdepodobnosť správneho a efektívneho nacvičenia neurónovej siete alebo ML modelu.

Táto situácia je pravdepodobne spôsobená odchýlkou 3D pose modelu ktorý predikoval výsledky nesprávne. Ďalším faktorom ktorý mohol spôsobiť tieto nekvalitné výsledky môžeme považovať kvalitu dátovej sady, kde snímky majú nízke rozlíšenie a póza nie je vždy celá viditeľná na snímku.

4.4 Trénovanie Machine Learning modelu

V tejto sekcii sú popísané experimenty ktoré boli vykonané pre tvorbu neurónových sietí a iných ML modelov pre účely klasifikácie identít a následnej analýze ich učenia a evaluácie. Experimenty zahŕňajú trénovanie neurónovej siete pre účely klasifikácie osôb. Pre tento účel boli zvolené existujúce metódy pre tvorbu neurónových sietí a klasických modelov strojového učenia (AutoML).

4.4.1 Trénovanie neurónovej siete

Ako typ použitej neurónovej siete bola zvolená jednoduchá neurónová sieť zložená z 3 vrstiev. Keďže vstupný vektor obsahuje 29 prvkov (podľa kostry *SMPL+head_30* 4.3.1) tak iný typ neurónovej siete (ako CNN) by sa zdal ako značne neefektívny.

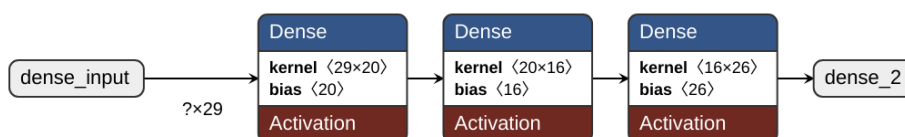
Na vytvorenie neurónovej siete bola použitá Python knižnica *Keras*² ktorá je nadstavbou ML knižnice *Tensorflow*. Pomocou Kerasu bola následne navrhnutá sieť s následujúcim zložením :

- **Vstupná vrstva** - veľkosť určuje veľkosť vektoru dĺžok
- **ReLU vrstva** - 20 neurónov
- **ReLU vrstva** - 16 neurónov
- **SoftMax vrstva** - počet neurónov je rovný počet vstupných tried - *výstupná vrstva*

Sieť je navrhnutá na viac-triednu klasifikáciu pričom boli využité nelineárne aktivačné funkcie pre možnosť učenia sa nad viac komplexnými dátami.

Ako vstup do neurónovej siete je možné predpokladať trénovacie dáta dĺžok bez označenia príslušnosti k triede. Táto hodnota je pri použití kostry *SMPL+head_30* 29 (30 kľúčových bodov kostry - 1 referenčná dĺžka *pelv-belly*).

Návrh ďalej obsahuje 2 skryté vrstvy ktoré majú aktivačnú funkciu ReLu. Tieto dve vrstvy majú za cieľ hľadať súvislosti medzi dátami. Ďalej sa v návrhu vyskytuje tretia, posledná vrstva ktorá zároveň slúži aj ako výstupná vrstva. Počet neurónov je priamo závislý od počtu unikátnych tried ktoré sa vo vstupných dátach nachádzajú. Táto vrstva ma za cieľ transformovať výstupy neurónov na pravdepodobnosti vďaka ktorým je možné predpovedať výstupnú triedu. Model následne vieme vizualizovať podľa obrázku 4.3.



Obr. 4.3: Neurónová sieť tvorená podľa návrhu 4.4.1 so 25 vstupnými triedami.

²Dostupné z : <https://keras.io/>

Trénovanie neurónovej siete na dátovej sade LaST a Player Reidentification Challenge

Na natrénovanie neurónovej siete bolo použitých prvých 25 identít z oboch dátových súrad. Dáta identít boli nezávislo od seba predspracované do tabuľkovej formy ktorá reprezentuje vzdialenosti medzi jednotlivými kľúčovými bodmi 4.3.2. Tieto dáta boli následne upravené pre potreby načítateľa dát (*Data Loader*) a rozdelené na tréningovú a testovaciu množinu v pomere 9:1.

Ako algoritmus pre optimalizátor (*optimizer*) bol zvolený **Adam (Adaptive Moment Estimation)**. Ako je možné vidieť aj na obrázku 4.2, v dátach sa často vyskytuje údaj *-1* ktorý značí že údaj sa na obrázku nevyskytoval. Chýbajúce dáta môžu spôsobiť neefektívne učenie modelu, čo v konečnom dôsledku rapidne znižuje jeho kvalitu. Prvou z možností ako sa vysporiadať s chýbajúcimi hodnotami je nájsť stratovú funkciu ktorá bude tieto chýbajúce údaje najlepšie znášať. Druhou z možností je chýbajúce dáta nahradiť priemerom alebo hodnotou 3σ ktorá by tieto chýbajúce dáta nahradila.

Z dôvodu toho že ide o neurónovú sieť na rozpoznávanie osôb, by sa pomocou druhej možnosti vytvoril ešte väčší šum v dátach spôsobený nesprávnou interpretáciou chýbajúcich dát ktoré danú osobu neopisujú. Pri prvej možnosti môžeme uvažovať o regresných stratových funkciách alebo o pravdepodobnostnej funkciách ako *Mean Squared Error (MSE)*, *Mean Absolute Error (MAE)*, *Huber*, *LogCosh*, *Categorical Crossentropy*. Porovnanie týchto stratových funkcií je možné vidieť v tabuľke 4.3.

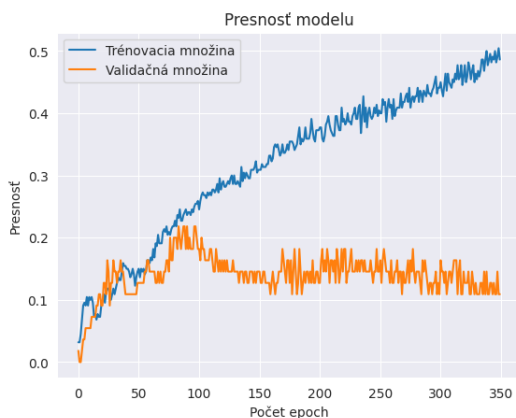
Tabuľka 4.3: Presnosť na validačnej množine pri použití rôznych stratových funkcií.

MSE	MAE	Huber	LogCosh	Categorical Crossentropy
0,0686	0,0784	0,0980	0,1275	0,2181

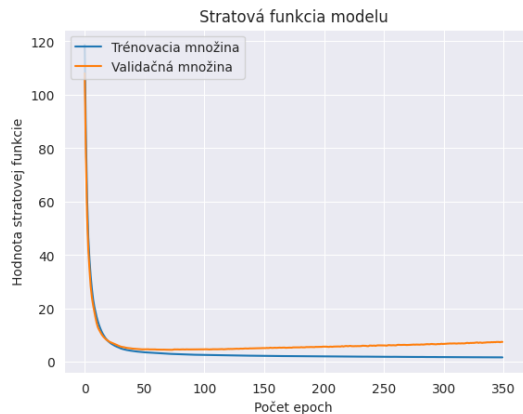
Ako je možné vidieť na tabuľke 4.3 najlepšie výsledky získaval model so stratovou funkciou **Categorical Crossentropy**. Keďže dát o jednej osobe/triede je málo, rozdelenie na tréningovú a validačnú množinu bolo nastavené na 8:2.

Pri procese učenia sa model cvičil po dobu 350 epoch kedy už bolo jasné že model nebude ďalej konvergovať. Podľa obrázku 4.4 je možné sledovať že model prestal konvergovať niekde okolo 100. epochy kedy je možné vidieť na krivke prepasovanie (*overfitting*) tým že krivka presnosti na validačnej množine začala klesať. Model pri tréningu skončil na presnosti **0.1028** na testovacej množine.

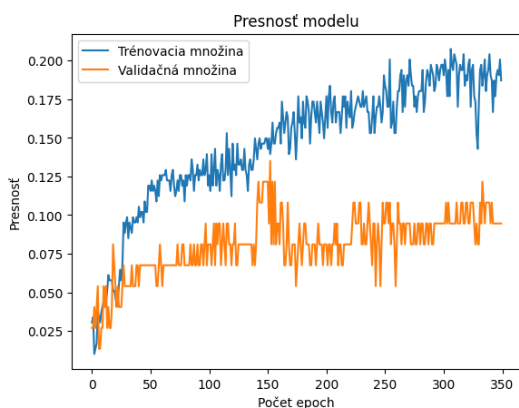
Pre proces učenia na dátovej sade **Player Reidentification Challenge** bol zvolený rovnaký postup ako pri predošlom učení. Na obrázku 4.6 je vidieť že model prestal konvergovať približne po 50 epoche kedy sa stala krivka presnosti na validačnej množine pomerne stabilná. Model po skončení tréningu dosahoval presnosť **0.0869** na testovacej množine.



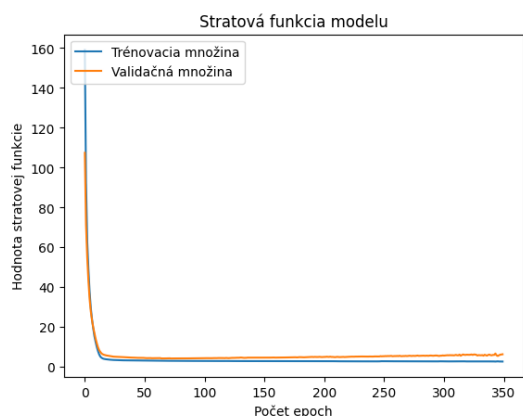
Obr. 4.4: Graf označujúci presnosť modelu počas učenia na datasete **LaST**.



Obr. 4.5: Graf označujúci hodnotu stratovej funkcie počas učenia na datasete **LaST**.



Obr. 4.6: Graf označujúci presnosť modelu počas učenia na datasete **Player Reidentification Challenge**.



Obr. 4.7: Graf označujúci hodnotu stratovej funkcie počas učenia na datasete **Player Reidentification Challenge**.

Vyhodnotenie

Výsledky z tréovania neurónových sietí pomocou oboch dvoch dátových súbív ukazujú, že modely neboli schopné dostatočne dobre zachytiť závislosti medzi jednotlivými triedami. Toto správanie je vysoko pravdepodobne spôsobené práve nízkou medzi-triednou variabilitou a vysokou vnútro-triednou variabilitou 4.3.3.

Výsledky, ktoré je možné pozorovať z tréovania na testovacej sade, ukazujú, že klasifikácia nie je dostatočne kvalitná a blíži sa k úrovni náhodnej klasifikácie. Z tohto dôvodu sa javí daný spôsob rozpoznávania ako *málo vhodný*.

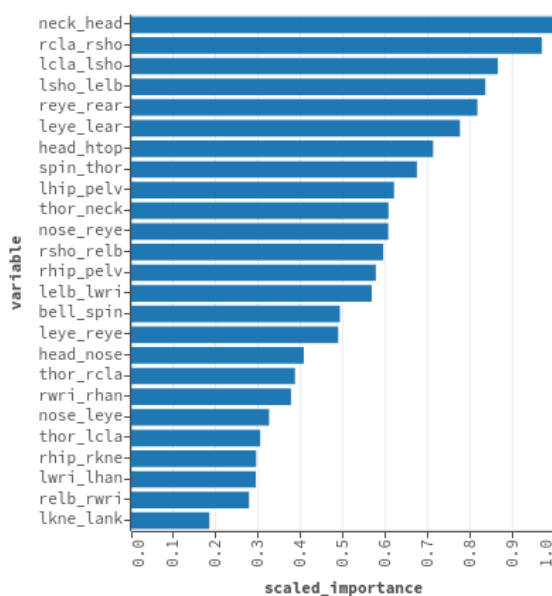
Ako ďalší poznatok, ktorý je možné z tréovania poznamenať, je rozdiel medzi tréovaním na dvoch dátových súbív. Pri cvičení na dátovej sade LAsT (viditeľné na obrázku 4.4) je oproti tréovaniu na druhom datasete (viditeľné na obrázku 4.6) možné pozorovať stabilnejšiu krivku učenia. To je pravdepodobne spôsobené vyššou kvalitou snímok u dátovej sady LAsT, aj pri skutočnosti, že LAsT obsahuje väčší priemerný počet chýbajúcich dĺžok v dátach.

4.4.2 Trénovanie ML modelu pomocou nástroja AutoML

V dátach ktoré máme k dispozícii často dochádza k tomu že atribúty nemajú danú hodnotu (nie sú viditeľné na snímku). Z toho dôvodu sa ponúka možnosť využiť iné druhy Strojového určenia na klasifikáciu ako je nástroj H2O AutoML³ ktorý dokáže vyprodukovať rôzne typy klasifikačných algoritmov ako *Random Forest*[30], *XGBoost*[30], *GBM*[30] a pod [30].

Postup prebiehal najprv definovaním typu dát a následným náhodným rozdelením dát k pomeru 8:1:1 na tréningovú, validačnú a testovaciu množinu. Nad danými množinami bol následne spustený AutoML pre tvorbu klasifikačných modelov ktorý začal s automatickým tréňovaním.

Ako najlepší model ktorý AutoML vyprodukovalo bol model **GBM_grid** (*Gradient Boosting Machine*). Tento model bol nacvičený s presnosťou **0.2745** na testovacej množine. Ako je možné vidieť obrázku 4.8, najdôležitejším atribútom pre model je vzdialenosť medzi krkom a vrchu hlavy⁴. Medzi atribútmi sa nenachádza žiaden razantný skok čo značí, že model nie je pre-učený.



Obr. 4.8: Dôležitosť/váha hodnoty atribútu na rozhodnutie modelu.

Vyhodnotenie

Výsledky ktoré vznikli pri tréňovaní sú síce porovnateľne lepšie oproti výsledkom z tréňovaní neurónových sietí popísaných v sekcii 4.4.1 ale stále sú na veľmi nízkej, v praxi nepoužiteľnej úrovni. Lepšie výsledky sú pravdepodobne spôsobené lepšou schopnosťou sa prispôbiť chýbajúcim hodnotám u klasický modelov strojového učenia. Tie sú schopné počas učenia tieto chýbajúce dáta úplne ignorovať a nezahrnúť ich do tréňovania bez nutnosti odstrániť celý záznam, ktorý obsahuje zvyšné atribúty.

³Dostupné z : <https://h2o.ai/platform/h2o-automl/>

⁴Atribút "Head" značí v kostre SMPL_30 vrch hlavy

Kapitola 5

Analytické a vizuálne rozpoznávanie osôb

Táto kapitola si berie za cieľ narozdiel od predchádzajúcej kapitoly, poňať úlohu analyticky a hľadať pre rozpoznávanie osôb algoritmické riešenie. Pre tento účel bude potreba nájsť tú najvhodnejšiu metódu, ktoré bude vedieť identity čo najlepšie od seba odlíšiť a tým pádom bude čo najefektívnejšie určovať identitu osoby na základe svojej znalostnej databázy.

V závere kapitoly bude za cieľ pripraviť riešenie ktoré preskočí proces extrahovania pózy a bude sa sústrediť čisto na vizuálnu podobnosť. Cieľom je zistiť, či táto metóda dokáže byť efektívna pre identifikáciu osôb a ak áno, aké výhody a nevýhody plynú s využitím tejto metódy.

5.1 Rozpoznávanie osôb na základe podobnosti vektorov

V minulej kapitole bola úloha poňatá spôsobom vytvorenia klasifikačného modelu ktorý musí byť zakaždým pretrénovaný pri pridaní ďalšej identity. Značnou nevýhodou bolo, že model nedokázal určiť a správne klasifikovať osobu nad ktorou nebol naučený a zakaždým sa ju snažil stotožniť s osobou ktorá bola obsiahnutá v tréningových dátach.

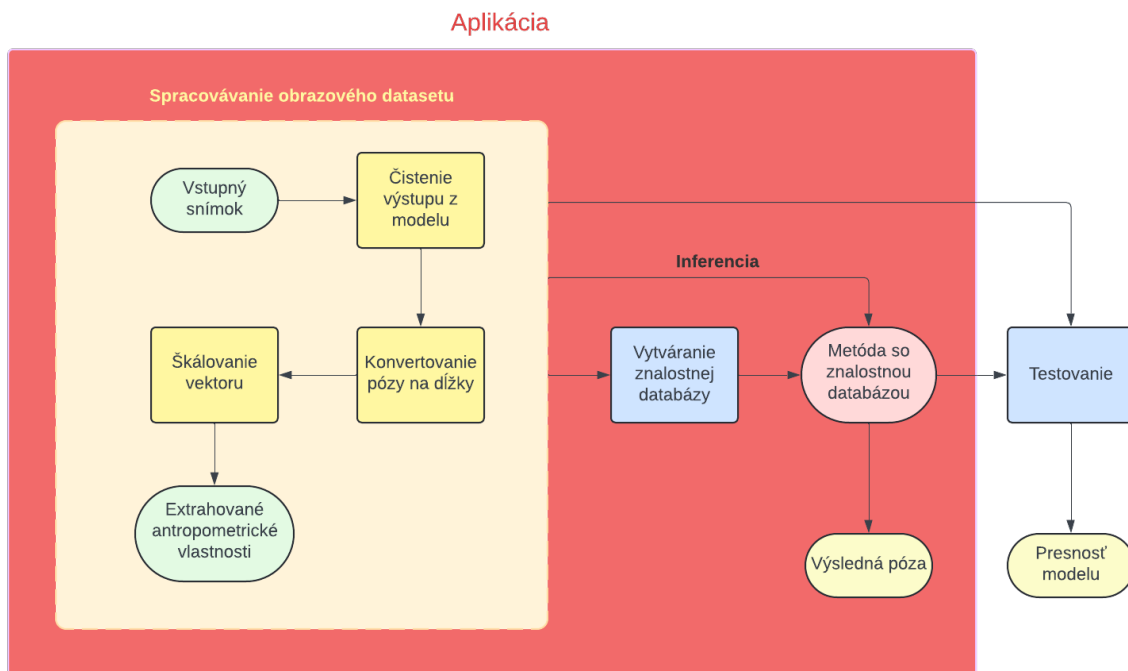
V tejto časti je za cieľ nájsť čo najefektívnejší spôsob ako osobu rozpoznať a stotožniť na základe rovnakých dát, ale spôsobom príbuznosti vektorov a rôznych metrík ktoré budú danú osobu v internej podobe popisovať.

5.1.1 Návrh riešenia

Extrakciu dát z obrazovej podoby na množinu dĺžok budeme praktikovať podobným spôsobom ako v minulej kapitole 4.1. Rozdielom bude vytváranie znalostnej databázy a následne určovanie výslednej identity. Znalostná databáza sa bude vytvárať podľa zvolenej metódy a rovnako tak určovanie výslednej identity.

Testovanie bude taktiež prebiehať rovnako ako v minulej kapitole 4.1, kde budeme považovať mieru úspešnosti modelu počet správne identifikovaných vzoriek ku celkovému počtu testovaných vzoriek.

Ďalším výsledkom každej metódy bude aplikácia ktorá bude tieto časti zaobalovať a dovoľovať jednoduché vytváranie znalostnej databázy a hľadanie najbližšej identity podľa zvolenej metódy. Vizuálnu podobu návrhu je možné pozorovať na obrázku 5.1.



Obr. 5.1: Návrh riešenia s využitím analytického rozpoznávania osôb.

5.1.2 Rozpoznávanie osôb podľa vzdialenosti vektorov

Vektor v ktorý vznikol pred-spracovaním dát z obrazovej podoby 4.3.2 obsahuje charakteristiku identity. Z toho dôvodu sa ponúka možnosť tieto vektory porovnávať a nájsť vektor/vektory zo znalostnej databázy ("trénovacej množiny") ktoré budú najbližšie popisovať pôvodný/vstupný vektor.

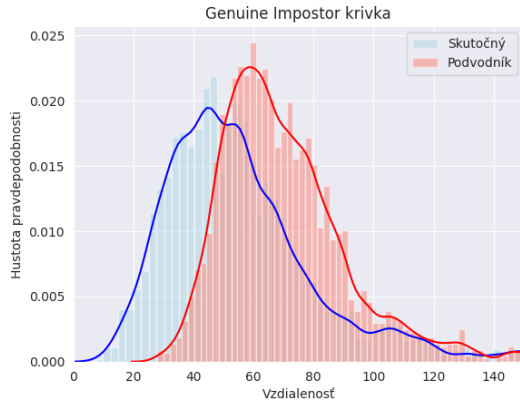
Ako metriku ktorá bude pre účely porovnávaní vektorov použitá bola zvolená **Euklidovská vzdialenosť vektorov** - d , ktorá v danom probléme bude definovaná nasledovne:

$$d = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (5.1)$$

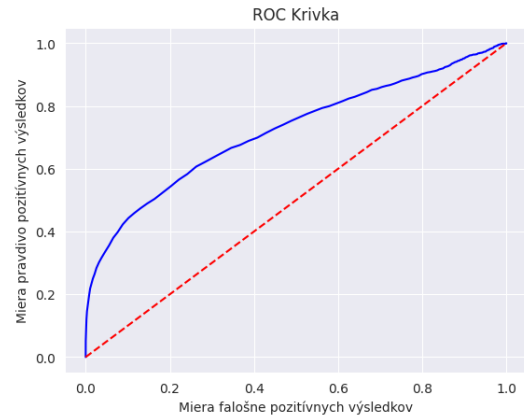
kde N je počet definovaných dĺžok vektorov, x je prvý vstupný vektor dĺžok a y je druhý vstupný vektor dĺžok.

Pri rozpoznávaní sa vypočíta Euklidovská vzdialenosť medzi vstupným vektorom a každým vektorom v znalostnej databáze. Tieto dĺžky sa následne zoradia od najmenej vzdialenosti a vyberie sa prvých k výsledkov ktoré značia výstupný vektor pre vstup. Ten sa ďalej prepočíta na základe zadanej konštanty k na výslednú triedu a presnosť s ktorou si je algoritmus istý že identita je správna. Ten je vypočítaný na základe počtu najčastejšie vyskytovanej identity vo výslednom k -rozmernom vektore, vydelený veľkosťou výstupného vektoru, ktorý je rovný k .

Vyhodnotenie metódy



Obr. 5.2: Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu podobnosti vektorov.



Obr. 5.3: Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu podobnosti vektorov.

Grafy 5.2 a 5.3 sú tvorené z pred-spracovaných dát z dátovej sady **Player Reidentification Challenge** z riadkov/záznamov ktoré obsahujú menej ako 3 chýbajúce hodnoty. Vzorka bola 25 identít s počtom celkových záznamov 460. Podľa výsledkov na grafe 5.2 je viditeľné že krivky oboch distribúcií sa značne prekrývajú, čo spôsobuje veľkú obtiažnosť nastaviť správny prah pre ktorý by bol použiteľný. Podľa grafu 5.2 a 5.3 by sa prahová hodnota mohla vyskytovať na úrovni vzdialenosti 40, ale aj v tomto prípade by dochádzalo ku častému nesprávnemu klasifikovaniu osoby.

Jednou z príčin, ktorá vedie k nevýhode tohto riešenia, je rozpoznávanie pri vstupe kde chýba väčší počet hodnôt. Ten istý prípad platí aj pri vytváraní znalostnej databázy kde dochádza k javu, ktorý spôsobuje zlú interpretáciu výsledkov. Tá vzniká pri počítaní euklidovskej vzdialenosti kedy sú chýbajúce hodnoty nahradené nulami, čo spôsobuje že vzdialenosť medzi takýmito hodnotami je rovná nule. V konečnom dôsledku to znamená, že vstupný vektor alebo vektor zo znalostnej databázy má väčšiu pravdepodobnosť že bude mať nižšiu euklidovskú vzdialenosť od iného vektoru (ktorý nemusí korešpondovať s rovnakou identitou) zo znalostnej databázy.

Ako ďalšiu nevýhodu ktorú je nutné uviesť pri tomto spôsobe riešenia je **časová zložitosť**. Pri rozpoznávaní je nutné porovnať a nájsť vzdialenosť vstupného vektoru s každým záznamom v znalostnej databáze, čo znamená že časová zložitosť tohto algoritmu je $O(n)$ kde n je počet všetkých prvkov v znalostnej databáze.

Naopak ako výhodu pri tomto spôsobe riešenia je možné určiť pomerne jednoduchú aktualizáciu znalostnej databázy. Algoritmus pri spustení nepotrebuje žiadnu ďalšiu úpravu dát, čo dovoľuje dáta rýchlo obmieňať bez nutnosti dlhšej inicializácie.

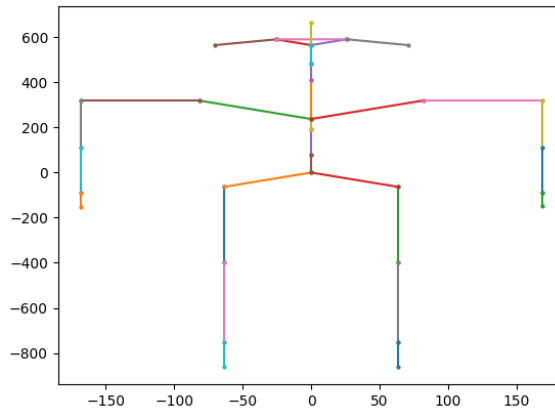
5.1.3 Rozpoznávanie osôb s použitím transformácie pózy

Ako ďalšia metóda ktorá bola nad dátami vyskúšaná pre úlohu rozpoznávania osôb je metóda, ktorej najväčší rozdiel oproti predchádzajúcim metódam spočíva v ďalšom pred-spracovaní dát. To spočíva vo forme prevodu existujúcich dát do predom určenej kostry, následnou transformáciou a výpočtu rozdielov medzi danými dvoma kostrami.

Tvorba referenčnej kostry

Referenčná kostra bude v tejto metóde slúžiť ako kostra, ku ktorej sa budeme musieť priblížiť počas transformácie. Táto kostra musí byť čo najviac všeobecná aby dokázala reprezentovať všetky kostry v znalostnej databáze a rozdelenie vzdialeností medzi nimi bolo korektne rozložené.

Pre tvorbu tejto kostry bol využitý priemer všetkých vzdialeností v danej kostre na základe ktorých bolo bod po bode kostra zložená. Kostra je zložená takým spôsobom, že body sú opísané 3D priestore so nulovou z-tovou súradnicou. Kostra ktorá následne pomocou postupu vznikla je vyobrazená na obrázku 5.4.



Obr. 5.4: Referenčná kostra vytvorená nad skráteným datasetom Player Reidentification Challenge.

Vytváranie znalostnej databázy

Ako vstupné dáta sú považované pre-procesované dáta ktoré vznikli podľa postupu v sekcii 4.3.2. Z každého záznamu sa následne vytvorí kostra podobne ako pri tvorbe referenčnej pózy v sekcii 5.1.3. Následne je nutná transformácia ktorá transformuje danú vstupnú kostru čo najbližšie k tej referenčnej.

Pre tieto účely transformácie bola použitá funkcia *Similarity Transform* z knižnice *skimage* (známa aj pod názvom *scikit-image*) ktorá využíva **škálovanie**, **rotáciu** a **transláciu** na transformáciu pôvodného vstupu (vstupnej kostry) čo najbližšie k vstupnej referenčnej kostre. Pri tejto transformácii nedochádza k žiadnej zmene tvaru pôvodnej kostry a tým pádom sémantika pózy zostáva nezmenená.

Metrika rozdielov póz

Pre porovnanie rozdielov póz bola zvolená metrika Stredná kvadratická chyba (*Mean squared error*). Ktorá je definovaná pre daný prípad následovne:

$$MSE = \frac{1}{N} \sum_{i=1}^N [(x_{i_1} - y_{i_1})^2 + (x_{i_2} - y_{i_2})^2 + (x_{i_3} - y_{i_3})^2] \quad (5.2)$$

kde N je počet všetkých definovaných bodov kostry, x je vstupný transformovaný vektor kostry a y je referenčný vektor kostry.

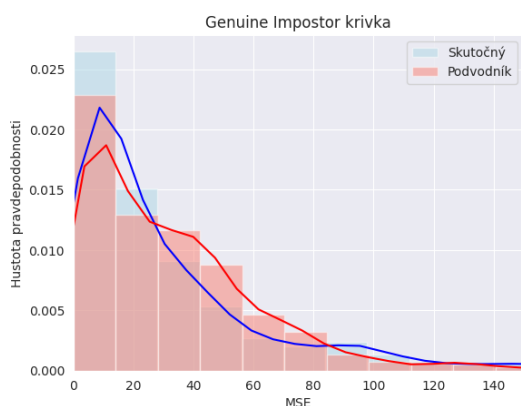
Metrika **MSE** bola zvolená práve na základe jednej z jej hlavných vlastností - *citlivosť na odľahlé hodnoty (kvadratická povaha)*. Tým že sa póza bude ďalej od priemeru, jej chyba sa bude kvadraticky navyšovať a tým pádom bude aj lepšie zachytiteľná a odlišiteľná. Táto vlastnosť by mala metóde pomôcť správne odlišiť triedy/identity v dátach a tým pomôcť ku celkovej výkonnosti.

Predikcia výslednej triedy

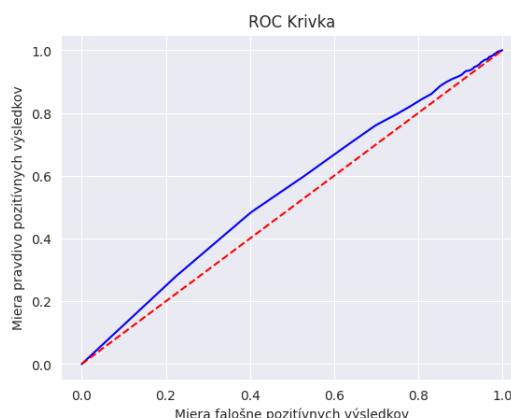
Znalostnú databázu, ktorá bola vytvorená v predchádzajúcom kroku v sekcii 5.1.3, budeme využívať na porovnávanie vstupných dát pre určenie konečnej triedy. Konkrétne pre tieto účely budeme využívať hodnotu MSE daného riadku.

Vstupný vektor, ktorý bol vytvorený pred-procesovným obrazových dát v sekcii 4.3.2 konvertujeme rovnakým spôsobom na kostru ako pri vytváraní znalostnej databázy v sekcii 5.1.3. Následne je cieľom, vybrať k záznamov zo znalostnej databázy ktoré budú mať čo najbližšiu absolútnu vzdialenosť MSE ku vstupnému vektoru. Podobne ako pri predchádzajúcej metóde podobnosti vektorov 5.1.2, zo zoznamu najbližších hodnôt MSE k vstupnému vektoru sa vyberie najviac vyskytovaná trieda. Počet výskytov konečnej triedy sa vydelí celkovou veľkosťou zoznamu najbližších prvkov čím vieme určiť precíznosť - ako veľmi si je metóda istá že vstupný vektor patrí výslednej identite.

Vyhodnotenie metódy



Obr. 5.5: Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu s transformáciou pózy.



Obr. 5.6: Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu s transformáciou pózy.

Grafy 5.5 a 5.6 sú rovnako ako pri vyhodnocovaní metódy podobnosti vektorov v sekcii 5.1.2 tvorené z pred-spracovaných dát z dátovej sady Player Reidentification Challenge z riadkov/záznamov ktoré obsahujú menej ako 3 chýbajúce hodnoty. Vzorka bola 25 identít s počtom celkových záznamov 460.

ROC krivka znázornená na obrázku 5.6 síce ukazuje že metóda dokáže kvalifikovať lepšie ako náhodný klasifikátor, ale rozdiel je veľmi malý. To potvrdzuje aj graf distribúcie podvodníka a autentickú distribúcie na obrázku 5.5 ktorý znázorňuje, že vzdialenosť medzi výslednými hodnotami MSE nie je signifikantná a je podľa nej komplikované určiť identitu osoby. Obe distribúcie na grafe 5.5 sa vo veľkej miere prelínajú čo značí že aj definovanie

prahovej hodnoty by bolo do veľkej miery komplikované.

Výhodou tejto metódy je rýchlosť, referenčná kostra ako aj znalostná databáza sa inicializujú na začiatku. Následne pri inferencii sa počíta MSE iba pre vstupnú kostru a hodnoty sa iba porovnávajú. Týmto spôsobom dokáže byť inferencia rýchlejšia.

Naopak jedná z hlavných nevýhod tejto metódy pochádza z predošlej výhody - čas inicializácie a aktualizácia znalostnej databázy. Pri aktualizácii je nutná aj výmena referenčnej kostry a to z toho dôvodu že je viazaná na priemerné hodnoty každého atribútu v dátach. Na základe tejto kostry sú vypočítavané MSE čo spôsobuje, že aj keby sa kostra vymenila, pôvodne vypočítané MSE by nemuseli zodpovedať k novo-vytvorenej kostre.

5.1.4 Rozpoznávanie osôb podľa chyby pózy

V predchádzajúcej metóde transformácie pózy v sekcii 5.1.3 je možné pozorovať pokles presnosti modelu oproti metóde porovnávania vzdialeností vektorov 5.1.2. Z toho dôvodu sa ponúka tieto dve metódy zlúčiť a navrhnúť metódu ktorá bude hľadať súvislosti medzi dvoma umelo vytvorenými kostrami.

Základom tejto metódy bude pre každý záznam v znalostnej databáze vytvoriť kostru podľa atribútov obsiahnutých v jednom zázname a následne vytvoriť kostru aj pre vstupný vektor. Tieto vektory bude nutné následne porovnávať a nájsť k vektorov, ktorých metrika bude čo najbližšie ku vstupnému vektoru.

Vytvorenie znalostnej databázy

Účelom znalostnej databázy bude, ako pri predchádzajúcich riešeniach, byť porovnávacou množinou s ktorými sa bude vstupný vektor porovnávať. Pri vytváraní kostier bol použitý rovnaký postup ako pri metóde transformácii pózy v sekcii 5.1.3 s tým rozdielom, že výsledná kostra sa neporovnáva so žiadnou referenčnou kostrou, ale kostra sa uloží ako atribút pre porovnávanie.

Použitá metrika pre porovnávanie

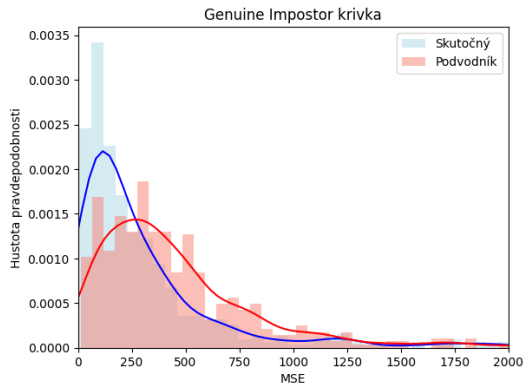
Ako metriku ktorá bude použitá pri hľadaní najbližšej kostry bola opäť použitá metrika MSE (Mean Squared Error). Využívame pri nej rovnakú vlastnosť ako pri predchádzajúcom riešení a to *citlivosť na odľahlé hodnoty*. Rozdiely medzi jednotlivými kostrami sa budú líšiť kvadraticky čo pomôže lepšie diverzifikovať rozdiely medzi jednotlivými identitami. Ďalšou jej výhodou ktorá bude využitá je jej vlastnosť - *pozitívnosť*. Tá značí že metrika bude vždy vracat iba kladné čísla a tým pádom bez ďalších úprav bude značiť rozdiel pózy.

Predikcia výslednej triedy

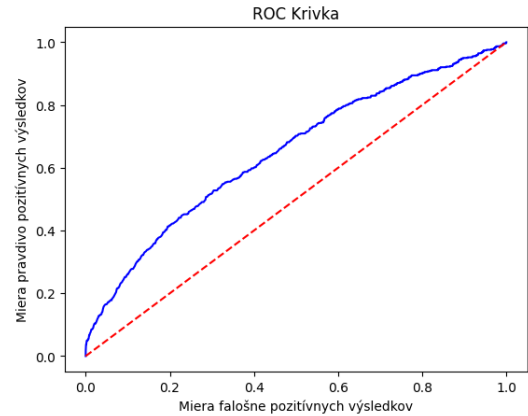
Vstupný vektor ktorý vznikol pred-procesovaním vstupných obrazových dát v sekcii 4.3.2 konvertujeme na predom určený typ pózy podľa sekcii 5.1.3 ktorý bude slúžiť ako jeden z argumentov pre porovnávanie. Následne porovnáme pózu s každou pózou ktorá je obsiahnutá v znalostnej databáze a vypočítame MSE pre každý záznam. Tieto hodnoty MSE zoradíme a vyberieme k najbližších hodnôt. Z týchto hodnôt vyberieme najčastejšie sa vyskytovanú triedu ktorú budeme považovať za výslednú a počet výskytov výslednej triedy

vydelíme veľkosťou zoznamu najbližších hodnôt - k . Táto hodnota nám bude vyjadrovať s akou presnosťou si je metóda výberom istá.

Vyhodnotenie metódy



Obr. 5.7: Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu chyby pózy.



Obr. 5.8: Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu chyby pózy.

Podľa ROC krivky na grafe 5.8 je možné konštatovať že systém má určitú schopnosť rozlišovať medzi pravdivými a podvodnými vzorkami. Distribúcia podvodníka a autentická distribúcia na grafe 5.7 je taktiež ako u predošlých metódach značne prekrytá ale miera prekrytia oproti metóde transformácie pózy 5.1.3 je značne menšia. Podľa grafu distribúcií 5.7 vieme taktiež určiť bod, kedy sa tieto dve distribúcie prelínajú a podľa neho nastaviť hranicu (*threshold*). Tieto distribúcie sa prvý krát prelínajú na hodnote okolo MSE 270. Pri zvolení tejto hranice je však nutné počítať s veľkým množstvom falošne pozitívnych výsledkov.

Táto metóda kombinuje výhody a nevýhody metódy vzdialenosti vektorov 5.1.2 a metódy transformácie pózy 5.1.3. Oproti metóde transformácie pózy 5.1.3, metóda nepotrebuje dlhý čas inicializácie pri ktorej by bolo potrebné znovu vypočítať novú hodnotu, tým pádom aj aktualizácia znalostnej databázy je relatívne rýchla. Metóda však vyžaduje dlhší čas pri inferencii a to z dôvodu toho, že kostra ktorá vznikla zo vstupného vektoru, musí byť porovnaná s každým prvkom v znalostnej databáze.

Oproti metóde vzdialenosti vektorov má metóda mierne pomalšiu inferenciu z dôvodu, že je nutné vypočítať MSE z trojdimenzionálnych bodov. Taktiež na časovú obtiažnosť pôsobí nutná konverzia do predom definovanej kostry.

5.2 Klasifikácia osôb na základe vizuálnej podobnosti

Ako ďalšou možnosťou ktorou je možné antropometrické vlastnosti zachytávať a následne porovnávať je pomocou **Embedding modelov**. Tieto modely dokážu konvertovať obrazovú podobu modelu na ich numerickú reprezentáciu. Táto numerická podoba obrazu reprezentuje vektory, ktoré zachytávajú na danom obrázku dôležité vlastnosti. Numerická reprezentácia snímok dovoľuje ďalším aplikáciám porozumieť čo sa na obrázku nachádza,

alebo v tomto prípade najdôležitejšie, ktoré snímky sú si najviac blízke. Týmto spôsobom vieme navrhnúť metódu ktorá dokáže snímky osôb previesť na numerickú metódu (*embedovať*), vytvorí vektorovú databázu a následne podľa vstupného obrazu nájst z vektorovej databázy najbližší vektor a tým určiť výslednú identitu osoby.

5.2.1 Výber detekčného modelu

Pre lepšiu presnosť embedding modelu je nutné mať na obraze iba osobu ktorú chceme pridať do databázy alebo podľa nej hľadáme najlepšiu zhodu. Pre tento účel bol použitý detektor osôb **YOLOv8**¹. YOLO 3.2.1 podporuje detekovanie 80 tried podľa datasetu COCO, ale pre účely detekcie osôb sa bude uvažovať iba trieda 'person' - osoba.

5.2.2 Metrika pre porovnávanie tried

Porovnávanie vektorov môže byť uskutočnené viacerými metrikami ako : *Euklidovská vzdialenosť medzi koncami vektorov*, *Podobnosť vektoru na základe vnútorného súčinu* alebo *kosínusová podobnosť*. Keďže pri porovnávaní nás hlavne zaujíma orientácia vektorov (ako moc sú vektory orientované na seba), bola zvolená metrika **kosínusová podobnosť** ktorá je definovaná nasledovne:

$$s = \frac{\vec{A} * \vec{B}}{\|\vec{A}\| \|\vec{B}\|} \quad (5.3)$$

kde \vec{A} je prvý vstupný vektor, \vec{B} je druhý vstupný vektor a $\|\vec{A}\|$ a $\|\vec{B}\|$ sú veľkosti vektorov.

Výsledok kosínusovej podobnosti je z intervalu $< -1, 1 >$ kde:

- 1 značí že vektory sú identické
- 0 značí žiadnu podobnosť medzi vektormi (*ortogonalitu*)
- -1 značí že dva vektory sú diametricky odlišné

Na základe tejto metriky vieme určiť ako blízko dané vektory sú tým, že kosínusová podobnosť musí byť čo najväčšia. Spolu s triedou najbližšieho vektoru vieme zistiť taktiež aj mieru podobnosti a tým nastaviť ideálnu prahovú hodnotu (*threshold*) pre systém založený na tejto metóde.

5.2.3 Výber embedding modelu

Ako *Embedding* model bol použitý model od Googlu ktorý je súčasťou Open-Source-ovej knižnice Mediapipe². Knižnica a model podporujú pre-procesovanie obrázku ako rotovanie obrazu, zmenu veľkosti alebo normalizáciu. Z toho dôvodu nie je potrebná žiadna ďalšia úprava a je možné snímok spracovať pomocou knižnice.

Taktiež pre výpočet kosínusovej podobnosti (popísané v sekcii 5.2.2) bude použitá funkcia *cosine_similarity* ktorá si berie dva argumenty - vektory, medzi ktorými podobnosť vypočíta.

¹Dostupné z : <https://docs.ultralytics.com/tasks/detect/>

²Dostupné z : <https://github.com/google/mediapipe>

5.2.4 Návrh systému

Návrh systému spočíva vo vytvorení znalostnej databázy ktorá bude slúžiť ako nástroj pomocou ktorého sa budú vstupné snímky porovnávať kosínusovou podobnosťou. Pre tento účel bude slúžiť vektorová databáza ktorá uchováva dvojicu vektor a triedu. Výsledkom predikcie by teda mala byť trieda a precíznosť podľa zvolenej konštanty k .

Vytvorenie znalostnej/vektorovej databázy

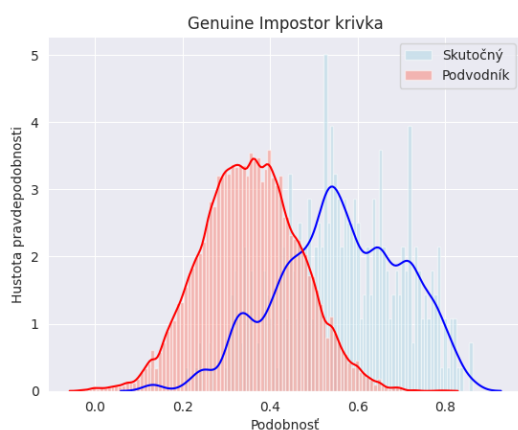
Pre každý snímok ktorý máme v dátach je nutné vytvoriť vektor pre porovnávanie. Táto konverzia prebieha najprv predikciou osoby na obrázku a vyrezaním všetkých ostatných rušivých elementov detekčným modelom (popísaný v sekcii 5.2.1). Snímok sa následne podľa predikcii z detektoru oreže podľa predpovedaného *bounded-boxu*. *Embedding* model vytvorí zo snímku vektor, ktorý sa uloží do predom pripraveného objektom aj s triedou/identitou ktorú vektor reprezentuje.

Predikcia výslednej pózy

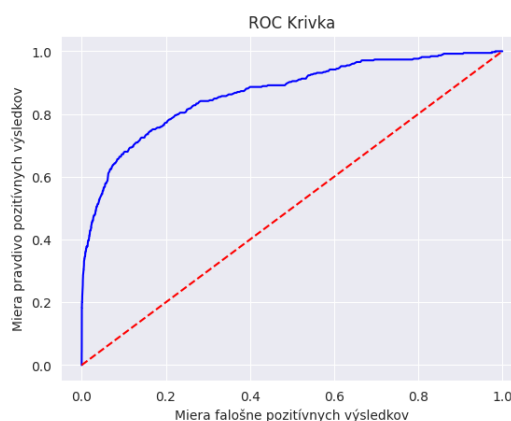
Pre vstupný snímok sú najskôr podobne ako pri vytváraní vektorovej databázy najprv predikované *bounded boxes*, z ktorých sa vyberie *bounded box* jednej osoby a snímok sa následne podľa neho oreže. Z orezaného snímku sa vytvorí pomocou *embedding* modelu jeho vektorová reprezentácia ktorá sa uloží na porovnávanie.

Pri hľadaní najbližšej zhody pre vektor vstupného snímku sa pre každý záznam v znalostnej databáze vypočíta kosínusová podobnosť s vektorom vstupného snímku. Po výpočte sa následne vyberie k záznamov s podobnosťou ktorá je najbližšie k úrovni 1. Z týchto hodnôt sa vyberie najčastejšie sa vyskytovaná trieda podľa ktorej sa vypočíta aj precíznosť, s akou istotou si je metóda istá že odpoveď je správna na základe konštanty k .

5.2.5 Vyhodnotenie metódy



Obr. 5.9: Graf znázorňujúci distribúciu podvodníka (červená) a autentickú distribúciu (modrá) pre metódu vizuálnej podobnosti.



Obr. 5.10: Graf ROC krivky (modrá) v porovnaní s náhodným klasifikátorom (oranžová) pre metódu vizuálnej podobnosti.

Pre overenie validity metódy bola vybraná skrátaná obrazová dátová sada Player Reidentification Challenge Dataset s 25 identitami a celkovým počtom 485 snímok. Z týchto snímok boli náhodne poskladané pozitívne a negatívne dvojice a medzi nimi bola vypočítaná miera kosínusovej podobnosti. Následne podľa týchto dvojíc boli vytvorené graf distribúcie podvodníka a autentická distribúcia 5.9 a graf ROC krivky (*Receiver Operating Characteristic curve*) 5.10.

V grafe autentickkej distribúcie a distribúcie podvodníka 5.9 môžeme pozorovať určité prekrytie týchto distribúcií. To značí že systém má určitú mieru chybovosti kedy môže označiť pozitívne pokusy za falošné a naopak. Prekrytie však nie je príliš významné a je možné pozorovať že distribúcia podvodníka v bode prieseku značne klesá. Ako prahovú hodnotu by sme mohli určiť pomocou grafov 5.9 a 5.10 na mieste okolo miesta kosínusovej podobnosti 0.5.

Výhodou tejto metódy sa zdá ako precíznosť na snímkoch testovacieho datasetu oproti predchádzajúcim metódam. Každopádne dátová sada často obsahuje snímky osôb ktoré boli zachytené v krátkom čase za sebou. To môže pôsobiť na výkonnosť systému tým spôsobom, že pokiaľ by sa v tréningovej množine vyskytovala napríklad osoba s jedným druhom oblečenia a vo validačnej iná identita s podobným druhom oblečenia, je vysoká pravdepodobnosť že by druhá osoba bola stotožnená s tou prvou s vysokým skóre podobnosti.

Ako nevýhodu môžeme ešte spomenúť dobu inicializácie, kedy sa vstupné snímky konvertujú na vektorovú databázu. Táto inicializácia vyžaduje načítanie modelu a *embedding* každého snímku čo môže byť, hlavne pri slabšom *hardware*, pomalá zalesnosť. Naopak výhoda je samotná inferencia, pri ktorej sa *embedduje* iba jeden snímok a vektor sa následne porovná s každým záznamom vo vektorovej databáze.

5.3 Súhrn uskutočnených experimentov

Pri analýze dát v sekcii 4.3.3 bolo zistené že vnútro triedna variabilita dát je znateľne väčšia ako medzi triedna variabilita. Táto situácia spôsobená estimátorom pózy popísaná v sekcii 4.3.1, do značnej miery spôsobuje horšie výsledky, ktoré boli zachytené ako výsledky testovania jednotlivých metód. Analýza medzitrjednej variability taktiež naznačuje že rozlíšenie tried je značne ťažšie, čo má za následok ďalšiu príčinu znižovania efektivity klasifikácie u metód.

V praxi je možné túto situáciu pozorovať pri tréningu Neurónových sietí v sekcii 4.4, kde je možné sledovať pri procese tréningu veľmi nízku presnosť. Tréningový proces, popísaný v sekcii 4.4.1, nedokázal jednoznačne odlíšiť tréningové dáta, pre účel klasifikácie osôb. Už pri počte 25 unikátnych osôb v testovacích dátach bola presnosť klasifikácie veľmi nízka a pridávaním ďalších identít by sme mohli pozorovať ešte nižšiu presnosť.

Podobné výsledky boli dosiahnuté aj pri tréningu "klasických" modelov strojového učenia pomocou AutoML v sekcii 4.4.2. Presnosť u týchto modelov bola mierne lepšia toto zlepšenie môžeme pripísať k lepšej schopnosti týchto modelov sa adaptovať na chýbajúce dáta. Dôležitú informáciu ktorú nám táto metóda poskytla bola metrika dôležitosti atribútov viditeľná na obrázku 4.8. Táto metrika ktorá značí dôležitosť atribútov pri rozhodovaní modelu, poskytuje znalosť, že na základe daných vlastností je možné rozpoznať osobu.

Ako ďalšie typy riešení, ktoré boli v tejto práci predstavené boli riešenia, ktoré boli založené na transformácii a porovnávaní. Prvým z týchto riešení bolo porovnávanie na základe vektorovej vzdialenosti pomocou euklidovskej vzdialenosti v sekcii 5.1.2. Táto metóda dokázala dáta klasifikovať čo najpresnejšie z metód, ktoré boli založené na podobnom riešení. To je možné sledovať na grafoch 5.3 a 5.2, kde sa dá pozorovať aspoň čiastočné oddelenie distribúcií u jednotlivých pozitívnych a negatívnych dvojíc.

Ďalším experimentom bolo vytvoriť metódu, ktorej cieľom má byť aspoň čiastočne eliminovať chybovosť pred spracúvaním dát v sekcii 5.1.3. Dáta sú v metóde prepočítané na kostru ktorá je následne pomocou funkcie *SimilarityTransform* transformovaná čo najbližšie k referenčnej kostre. Ako výsledok výpočtu sa považuje chyba voči referenčnej kostre. Metóda sa ukázala ako neefektívna, čo je možné sledovať na grafe distribúcie podvodníka a autentickej distribúcií 5.5. Je možné sledovať že metóda nemá skoro žiadnu schopnosť správne rozlíšiť pozitívne dvojice od negatívnych čo je možné taktiež pozorovať na obrázku ROC krivky 5.6.

Ako posledný experiment ktorý bol zvolený pre tento typ riešenia je porovnávanie podľa chyby pózy v sekcii 5.1.4. Táto metóda kombinuje postup pri porovnávaní vektorov 5.1.2 a postup pri rozpoznávaní osôb pomocou transformácie pózy 5.1.3. Pri tejto metóde sa podobne ako pri transformácii vytvorí kostra, ktorá je porovnávaná s každou kostrou v znalostnej databáze podobne ako pri porovnávaní vektorov. Tento postup dosahuje lepšie výsledky ako pri transformovaní pózy, čo je možné sledovať na obrázku ROC krivky 5.8 ale v porovnaní s metódou vzdialenosti vektorov sú výsledky znateľne horšie.

Možnosťou ako obísť pri rozpoznávaní estimátor pózy bolo riešenie pomocou *embedding* modelu 5.2. Pri tejto metóde sa pomocou detekčného modelu (v našom prípade model YOLO 5.2.1 vyreže osoba zo snímku a následne sa snímok pomocou *embedding* modelu (v našom prípade model v knižnici Mediapipe) prevedie na numerickú reprezentáciu. Takýmto spôsobom sa vytvorí nad všetkými tréningovými snímkami znalostná / vektorová databáza nad ktorou sa vstupný snímok porovnáva kosínusovou vzdialenosťou. Táto metóda sa vyskytla ako najefektívnejšia z predom spomenutých metód čo je možné sledovať aj na základe obrázku ROC krivky 5.10. Na grafe distribúcie podvodníka a autentickej distribúcií 5.9 je taktiež možné sledovať najmenšie prekrytie týchto distribúcií medzi sebou, čo značí že metóda má najväčšiu schopnosť rozlíšiť pozitívne vzorky od tých negatívnych. Na druhú stranu, ako bolo spomenuté vo vyhodnotení danej metódy 5.2.5, metóda prestáva byť účinná vplyvom času a podmienok. To je možné sledovať napríklad pri snímkoch jednej osoby v rôznom type oblečení zachytených v rôznych podmienkach a uhloch.

5.3.1 Súhrnné testovanie metód

Záverečné testovanie všetkých spomínaných metód spočíva v spustení každého skriptu v režime *validate*. Testovacia dátová sada bola vytvorená z podmnožiny 25 unikátnych osôb z originálneho datasetu Player Reidentification Challenge. Z každej triedy/identity v dátovej sade boli snímky rozdelené v pomere 8:2 na tréningovú a testovaciu množinu. Výsledky je možné pozorovať v tabuľke 5.1 kde je možné sledovať presnosť³ na testovacej dátovej sade pri použití rôznej konštanty k .

Na tabuľke 5.1 je možné pozorovať závery ktoré boli dosiahnuté v úvode tejto sekcie. Najlepšie výsledky dosahuje metóda Vizuálnej podobnosti 5.2 ktorá značne vedie v presnosti

³Presnosť v danom ponímaní rozumieme ako počet správne klasifikovaných záznamov ku počtu všetkých záznamov

Tabuľka 5.1: Presnosť jednotlivých metód pri rôznej hodnote konštanty k na testovacej vzorke.

k	Podobnosť vektorov	Transformácia pózy	Chyba pózy	Vizuálna podobnosť
1	0,396	0,099	0,319	0,875
3	0,374	0,110	0,242	0,865
4	0,385	0,099	0,242	0,885

na testovacej množine v porovnaní s ostatnými metódami. Naopak ako najmenej vhodná metóda sa ukázala metóda Transformácie pózy 5.1.3 ktorá sa javí pre toto riešenie ako najmenej presná.

Zmena konštanty k mala za dôsledok u väčšine metód zníženie presnosti. Túto situáciu môžeme odôvodniť viacerými faktormi ako nízky počet snímok v trénovacej množine alebo veľká odchýlka v odhadovaní pózy estimátorom. V ideálnom prípade by sa presnosť nemala meniť do hodnoty konštanty k , ktorú by predstavoval počet záznamov s najmenším počtom výskytu triedy. V testovacom prípade 5.1 bola hodnota počtu záznamov s najmenším počtom výskytu triedy 9. To zodpovedá zisteniam z jednotlivých metód zistených na začiatku sekcie a tomu že úroveň istoty metódy pri určovaní identity osoby je relatívne nízka.

Kapitola 6

Záver

V závere tejto práce môžeme konštatovať, že antropometrické vlastnosti majú obrovský potenciál pre možnosti identifikácie osoby. Kombinácie týchto vlastností plnia účel jedinečnosti čo ich robí výborným kandidátom pre identifikáciu osoby. Podrobná analýza antropometrických vlastností taktiež ukázala že tieto vlastnosti sa síce časom menia, ale vývin vlastností je dostatočne pomalý aby bolo možné detekovať osobu aj s odstupom času.

Práca taktiež preskúmala existujúce riešenia rozpoznávania osôb ktoré sú široko používané vo viacerých oblastiach. Tieto riešenia boli porovnávané, pričom bol dôraz na poukázanie ich výhod a nevýhod, a taktiež ich spôsob využitia v reálnom svete. Pri rozbore existujúcich riešení (odtlačky prsta, oblasť tváre) pre identifikáciu je možné konštatovať že antropometrické vlastnosti majú potenciál pre využitie hlavne v podmienkach kde nie je možné využiť existujúce prostriedky a metódy. Napríklad pre odvrátenú tvár od kamery alebo skryté odtlačky prstov. Pri experimentovaní sa však ukázalo, že existujúce riešenia pre odhad pózy, nedokážu dostatočne presne odhadnúť pózu osoby pre účely následnej identifikácie. To spôsobovalo veľké problémy pri testovaní metód práce, keďže vlastnosti ktoré boli pomocou modelu pre odhad pózy extrahované, mali pri jednotlivých identitách vysoké odchylky.

Ďalším zistením bola nevhodnosť využitia neurónových sietí pre účely klasifikovania osôb. Toto riešenie sa potýkalo s podobnými problémami ako pri riešeníach, ktoré strojové učenie nepoužívali. Medzi tieto problémy patrí hlavne vysoká triedna a nízka medzitriedna variabilita. Ďalšie problémy pre strojové učenie spôsobovali aj chýbajúce dáta, s ktorými sa lepšie vedeli vysporiadať ostatné metódy.

Výzvy s ktorými sa pracovalo v riešení s neurónovými sieťami bolo možné pozorovať aj u analytických riešení ktoré síce dosahovali všeobecnejšie lepšej presnosti ale stále silno trpeli veľkou variabilitou dát ktorá znemožňovala metódam správne vstupný vektor zaradiť ku vektorom zo znalostnej databázy a tým aj správne určiť výslednú identitu.

Zásadné zistenie prinieslo riešenie pomocou vizuálnej podobnosti ktoré, oproti ostatným metódam, vedelo vysoko kvalitne predikovať identity osôb. Zároveň, toto riešenie aj pri svojich nevýhodách pôsobí ako aktuálne najvhodnejšie riešenie pre rozpoznávanie osôb na základe antropometrických vlastností.

Zároveň má zmysel podotknúť že riešenia ktoré boli založené na extrahovaní pózy pomocou estimátoru pózy trpeli vysokou variabilitou. Do budúca sa tým pádom ponúka možnosť tieto metódy testovať s novými modelmi ktoré budú mať vyššiu precíznosť a budú

vedieť kvalitnejšie a presnejšie extrahovať pózu z obrazu. To môže viesť k lepším výsledkom so zachovaním výhod, ktoré rozpoznávanie pomocou antropometrických vlastností do sveta biometrických systémov prináša.

Literatúra

- [1] AGUILAR, G., SANCHEZ, G., TOSCANO, K., SALINAS, M., NAKANO, M. et al. Fingerprint Recognition. 2007, s. 32–32, [cit. 2024-01-04]. DOI: 10.1109/ICIMP.2007.18. Dostupné z: <https://ieeexplore.ieee.org/document/4271777>.
- [2] ALBAWI, S., MOHAMMED, T. A. a AL ZAWI, S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering and Technology (ICET)*. 2017 [cit. 2023-12-06]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/8308186>.
- [3] ALI, M., MAHALE, V., YANNAWAR, P. a GAIKWAD, A. Overview of Fingerprint Recognition System. Marec 2016, [cit. 2023-01-04]. DOI: 10.1109/ICEEOT.2016.7754902. Dostupné z: www.researchgate.net/publication/310953762_Overview_of_Fingerprint_Recognition_System.
- [4] CASADEI, K. a KIEL, J. *Anthropometric Measurement*. 2023 [cit. 2023-12-18]. [Updated 2022 Sep 26]. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK537315/>.
- [5] DELAC, K. a GRGIC, M. A survey of biometric recognition methods. 2004, s. 184–193, [cit. 2024-01-03]. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/1356372>.
- [6] FAUNDEZ ZANUY, M. Biometric security technology. *Aerospace and Electronic Systems Magazine, IEEE*. Júl 2006, zv. 21, s. 15 – 26, [cit. 2023-12-08]. DOI: 10.1109/MAES.2006.1662038.
- [7] FERRAZ, R., BRANQUINHO, L., LOUPO, R., NEIVA, H. a MARINHO, D. The relationship between anthropometric characteristics and sports performance in national-level young swimmers. *European Journal of Human Movement*. 2020, [cit. 2023-12-18]. DOI: 10.21134/EURJHM.2020.45.2.
- [8] GLENN, J. *Pose estimation datasets overview*. Nov 2023 [cit. 2024-02-27]. Dostupné z: <https://docs.ultralytics.com/datasets/pose/#usage>.
- [9] HANGARAGI, S., SINGH, T. a N, N. Face Detection and Recognition Using Face Mesh and Deep Neural Network. *Procedia Computer Science*. 2023, zv. 218, s. 741–749, [cit. 2024-01-03]. DOI: <https://doi.org/10.1016/j.procs.2023.01.054>. ISSN 1877-0509. International Conference on Machine Learning and Data Engineering. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1877050923000546>.

- [10] HEALTH STATISTICS, N. C. for. *Growth charts - clinical growth charts*. Centers for Disease Control and Prevention, Jun 2017 [cit. 2023-11-12]. Dostupné z: https://www.cdc.gov/growthcharts/clinical_charts.htm.
- [11] HSIAO, H., SNYDER, K. a LONG, D. *Anthropometry*. Centers for Disease Control and Prevention, Aug 2022 [cit. 2023-12-31]. Dostupné z: <https://www.cdc.gov/niosh/topics/anthropometry/default.html>.
- [12] JORDAN, M. I. a MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*. 2015, zv. 349, č. 6245, s. 255–260, [cit. 2023-01-12]. DOI: 10.1126/science.aaa8415. Dostupné z: <https://www.science.org/doi/abs/10.1126/science.aaa8415>.
- [13] KENDALL, A., GRIMES, M. a CIPOLLA, R. *PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization*. 2016 [cit. 2024-02-28]. Dostupné z: <https://arxiv.org/abs/1505.07427>.
- [14] KRISHAN, K. Anthropometry in Forensic Medicine and Forensic Science- 'Forensic Anthropometry'. 2018, [cit. 2023-10-12]. DOI: 10.5580/1dce.
- [15] LAJČI, U. a MISINI, E. Biometric Authentication. December 2022, [cit. 2023-11-15].
- [16] LUXIMON, A. a ZHANG, M. 6 - The human body. In: LI, Y. a DAI, X.-Q., ed. *Biomechanical Engineering of Textiles and Clothing*. Woodhead Publishing, 2006, s. 91–110 [cit. 2023-11-10]. Woodhead Publishing Series in Textiles. DOI: <https://doi.org/10.1533/9781845691486.2.91>. ISBN 978-1-84569-052-6. Dostupné z: <https://www.sciencedirect.com/science/article/pii/B9781845690526500066>.
- [17] MAJI, D., NAGORI, S., MATHEW, M. a PODDAR, D. *YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss*. 2022 [cit. 2024-03-01]. Dostupné z: <https://arxiv.org/ftp/arxiv/papers/2204/2204.06806.pdf>.
- [18] NOORWALI, E. A. a ALJAADI, A. M. A Review of Anthropometric Measurements for Saudi Adults and Elderly, Directions for Future Work and Recommendations to Establish Saudi Guidelines in Line with the Saudi 2030 Vision. *Healthcare*. 2023, zv. 11, č. 14, [cit. 2023-12-20]. ISSN 2227-9032. Dostupné z: <https://www.mdpi.com/2227-9032/11/14/1982>.
- [19] OFEIDIS, I., KIEDANSKI, D. a TASSIULAS, L. *An Overview of the Data-Loader Landscape: Comparative Performance Analysis*. 2022 [cit. 2024-02-26]. Dostupné z: <https://arxiv.org/pdf/2209.13705.pdf>.
- [20] PROF.A.D. DONGARE, P. K. A. K. Introduction to Artificial Neural Network. *International Journal of Engineering and Innovative Technology (IJEIT)*. 2012, [cit. 2024-02-10]. Dostupné z: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=04d0b6952a4f0c7203577afc9476c2fcab2cba06>.
- [21] RAHULKISHOREBDM. *Machine Learning in Business Intelligence*. Medium, Mar 2020 [cit. 2023-12-01]. Dostupné z: <https://medium.com/@rahulkishorebdm/machine-learning-in-business-intelligence-6ed6980f37f9>.

- [22] RAJU, A. S. a UDAYASHANKARA, V. Biometric person authentication: A review. 2014, s. 575–580, [cit. 2024-01-03]. DOI: 10.1109/IC3I.2014.7019771. Dostupné z: <https://ieeexplore.ieee.org/document/7019771>.
- [23] SANTOS, D., DAWSON, J., MATIAS, C., ROCHA, P., MINDERICO, C. et al. Reference values for body composition and anthropometric measurements in athletes. *PLoS One*. Public Library of Science. May 15 2014, zv. 9, č. 5, s. e97846, [cit. 2023-12-20]. DOI: 10.1371/journal.pone.0097846. Dostupné z: <https://pubmed.ncbi.nlm.nih.gov/24830292/>.
- [24] SÁRÁNDI, I., HERMANS, A. a LEIBE, B. Learning 3D Human Pose Estimation from Dozens of Datasets using a Geometry-Aware Autoencoder to Bridge Between Skeleton Formats. In: *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023 [cit. 2023-12-08]. Dostupné z: <http://www.cstug.cz>.
- [25] SÁRÁNDI, I., LINDER, T., ARRAS, K. O. a LEIBE, B. MeTRAbs: Metric-Scale Truncation-Robust Heatmaps for Absolute 3D Human Pose Estimation. *IEEE Transactions on Biometrics, Behavior, and Identity Science*. 2021, zv. 3, č. 1, s. 16–30, [cit. 2024-02-11]. DOI: 10.1109/TBIOM.2020.3037257.
- [26] SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2021, zv. 2, č. 3, s. 160, [cit. 2024-01-05]. ISSN 2661-8907. Dostupné z: <https://doi.org/10.1007/s42979-021-00592-x>.
- [27] SHU, X., WANG, X., ZANG, X., ZHANG, S., CHEN, Y. et al. Large-Scale Spatio-Temporal Person Re-identification: Algorithm and Benchmark. *IEEE Transactions on Circuits and Systems for Video Technology*. 2022, s. 4390–4403, [cit. 2024-01-28].
- [28] SINGH, S. Anthropometric Measurements and Body Composition Parameters of Farm Women in North Gujarat. *Journal of Ergonomics*. Január 2013, zv. 03, [cit. 2023-12-08]. DOI: 10.4172/2165-7556.1000114. Dostupné z: www.researchgate.net/.
- [29] SOHAIL, A. S. M. a BHATTACHARYA, P. *Detection of Facial Feature Points Using Anthropometric Face Model*. Január 2008 [cit. 2023-12-20]. 189-200 s. ISBN 978-0-387-72499-7. Dostupné z: www.researchgate.net/publication/226925950_Detection_of_Facial_Feature_Points_Using_Anthropometric_Face_Model.
- [30] TAN, E. *Automated machine learning with H2O*. Towards Data Science, Aug 2021 [cit. 2024-03-01]. Dostupné z: <https://towardsdatascience.com/automated-machine-learning-with-h2o-258a2f3a203f>.
- [31] UTKUALP, N. a ERCAN, I. Anthropometric Measurements Usage in Medical Sciences. *Biomed Research International*. 2015, zv. 2015, s. 404261, [cit. 2023-12-31]. DOI: 10.1155/2015/404261. Epub 2015 Aug 27. Dostupné z: <https://doi.org/10.1155/2015/404261>.
- [32] WANG, J., TAN, S., ZHEN, X., XU, S., ZHENG, F. et al. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding*. 2021, zv. 210, s. 103225, [cit. 2023-12-18]. DOI: <https://doi.org/10.1016/j.cviu.2021.103225>. ISSN 1077-3142. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1077314221000692>.

- [33] ZANDYCKE, G. V., SOMERS, V., ISTASSE, M., DON, C. D. a ZAMBRANO, D. DeepSportradar-v1: Computer Vision Dataset for Sports Understanding with High Quality Annotations. In: *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*. ACM, Oct 2022 [cit. 2024-01-28]. DOI: 10.1145/3552437.3555699. Dostupné z: <https://doi.org/10.1145/3552437.3555699>.
- [34] ZOU, Z., CHEN, K., SHI, Z., GUO, Y. a YE, J. *Object Detection in 20 Years: A Survey*. 2023 [cit. 2024-03-10].
- [35] ZVONAR, M., ŠTEFAN, L., KASOVIĆ, M. a PILER, P. Tracking of anthropometric characteristics from childhood to adolescence: an 8-year follow-up findings from the Czech ELSPAC study. *BMC Public Health*. Apr 12 2022, zv. 22, č. 1, s. 727, [cit. 2023-12-20]. DOI: 10.1186/s12889-022-13178-w. ISSN 1471-2458. Dostupné z: <https://doi.org/10.1186/s12889-022-13178-w>.

Príloha A

Obsah pamäťového média

- **applications/** - aplikačná časť riešenia
- **research/** - výskumná časť riešenia
- **tex/** - zdrojové Latex súbory práce
- **requirements.txt** - potrebné knižnice pre fungovanie aplikačnej a výskumnej časti
- **technická_správa.pdf** - technická správa