

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

Ing. Martin Rajnoha

**URČOVANIE PODOBNOSTI OBJEKTŮ NA
ZÁKLADE OBRAZOVEJ INFORMÁCIE**

DETERMINATION OF OBJECTS SIMILARITY BASED ON
IMAGE INFORMATION

SKRÁTENÁ VERZIA DIZERTAČNEJ PRÁCE

Odbor: Teleinformatika

Školiteľ: doc. Ing. Radim Burget Ph.D.

Oponenti:

Dátum obhajoby:

KLÚČOVÉ SLOVÁ

superrozlíšenie, viacsnímkové, sekvencia, dataset, rozpoznávanie tvári, obraz-obraz, hlboké učenie, zvyšovanie kvality, rekonštrukcia obrazu, podobnosť obrazov

KEYWORDS

super-resolution, multi-frame, sequence, dataset, face-recognition, image-to-image, deep learning, quality increasing, image reconstruction, image similarity

MIESTO ULOŽENIA PRÁCE

Dizertačná práca je k dispozícii na Vedeckom oddelení dekanátu FEKT VUT v Brne, Technická 10, Brno, 616 00

© Rajnoha Martin, 2020

ISBN 80-214-

ISSN 1213-4198

Obsah

Úvod	4
1 Ciele dizertačnej práce	5
2 Navrhnuté metódy riešenia	6
2.1 Spracovanie obrazov pomocou tradičných metód	7
2.1.1 Rozpoznávanie ručne písaného textu	7
2.1.2 Rozpoznávanie materiálu zo spektroskopie	8
2.2 Podobnosť obrazov s použitím hlbokého učenia	8
2.2.1 Systém pre biometrickú identifikáciu ľudí	11
2.3 Zvyšovanie kvality snímku tváre	15
2.4 Zvyšovanie rozlíšenia zo sekvencie snímok	16
2.4.1 Návrh datasetu	17
2.4.2 Výber vhodných architektúr	18
2.4.3 Hodnotenie kvality	19
2.4.4 Návrh testu pre verejnosť	20
3 Overenie navrhnutých metód	21
3.1 Tradičné metódy	21
3.1.1 Rozpoznávanie ručne písaného textu	21
3.1.2 Rozpoznávanie materiálu zo spektroskopie	21
3.2 Hlboké učenie	22
3.3 Systém pre biometrickú identifikáciu ľudí	24
3.3.1 Detekcia príznakov pre Parkinsonovu chorobu	25
3.3.2 Vplyv pozadia na tréňovanie neurónovej siete	25
3.4 Jednosnímkové spracovanie	26
3.4.1 Zvyšovanie kvality tvárí	26
3.5 Spracovanie sekvencie snímok	28
3.5.1 Dataset	28
3.5.2 Porovnanie metód	29
3.5.3 Celosvetové porovnanie výsledkov	31
4 Záver	32
Literatúra	33
Curriculum Vitae	38

Úvod

Počítačové videnie je jedno z rýchlo sa rozvíjajúcich odvetví, s ktorým sa v živote bežne stretávame, a to v rôznych oblastiach. Pravdepodobne jednou z najvýznamnejších aplikácií merania podobnosti objektov na základe ich obrazovej informácie je rozpoznávanie osôb. Nakoľko je každodenný život ovplyvnený rôznymi hrozbami (terorizmus, krádeže, lúpežné prepadnutia...), veľké množstvo týchto aplikácií požaduje beh v reálnom čase, predovšetkým bezpečnostné systémy. S rastúcim technologickým pokrokom sa síce zlepšujú aj monitorovacie systémy, problémom však stále zostáva automatické spracovanie nekvalitných, resp. rozmazaných, alebo vzdialených objektov.

Táto práca sa zaoberá problémom rozpoznávania ľudí z nekvalitných záznamov a hľadá spôsob, akým vylepšiť súčasný stav poznania v tejto oblasti. Pred vymedzením sa na túto konkrétnu oblasť však bolo nutné vykonať ďalšie potrebné experimenty, ktoré stavajú na rovnakých základoch. Tieto experimenty, ktoré priniesli potrebný prehľad a zručnosti v danom odvetví, sú taktiež postupne popísané v tejto práci spoločne s dosiahnutými výsledkami. Tieto výskumné práce sa týkajú obrazového spracovania dát a začínajú základom do strojového učenia pomocou tradičných metód, neskôr prechádzajú na problematiku hlbokého učenia (klasifikácia, detekcia a podobnosť). Ďalej je vytvorený celý systém pre biometrickú identifikáciu, ktorý je následne vylepšený o superrozlíšenie z jedného snímku osoby. Nakoniec sa práca dostáva k jej hlavnému cieľu, ktorým je superrozlíšenie tváre z niekoľkých snímkov.

Hlavným prínosom práce je návrh metód pre superrozlíšenie tváre zo sekvencie snímkov nízkeho rozlíšenia. Jedná sa primárne o prípady, kde identifikácia osôb je na hranici rozpoznania, pričom navrhnuté metódy sa vymedzujú proti použitiu generatívneho prístupu (GANs). Pre overenie funkčnosti navrhnutých metód bola vytvorená nová tréningová a testovacia množina obsahujúca 17426 sekvencií tvárí nízkeho rozlíšenia. Veľkosťou je porovnateľná s poprednými svetovými databázami, avšak oproti stávajúcim databázam neobsahuje iba známe osobnosti, ale zahŕňa rôznych ľudí, rôzne etniká, obe pohlavia a širokú škálu veku, a obsahuje snímky z odlišných snímacích zariadení, nasvietení, uhlov apod. Metódy boli porovnávané na základe objektívnych metrík, ale aj pomocou ľudského hodnotenia, pričom bolo dokázané, že výsledky objektívnych metrík neodpovedajú vnímaniu človeka. Navrhnuté metódy prekonal výkonosť jednosnímkových metód, ktoré sú stále považované za najlepšie v súčasnosti, čím sa potvrdila hypotéza, že viacsnímkové metódy majú lepšiu schopnosť kvalitnejšie vytvoriť výstup ako jednosnímkové metódy.

1 Ciele dizertačnej práce

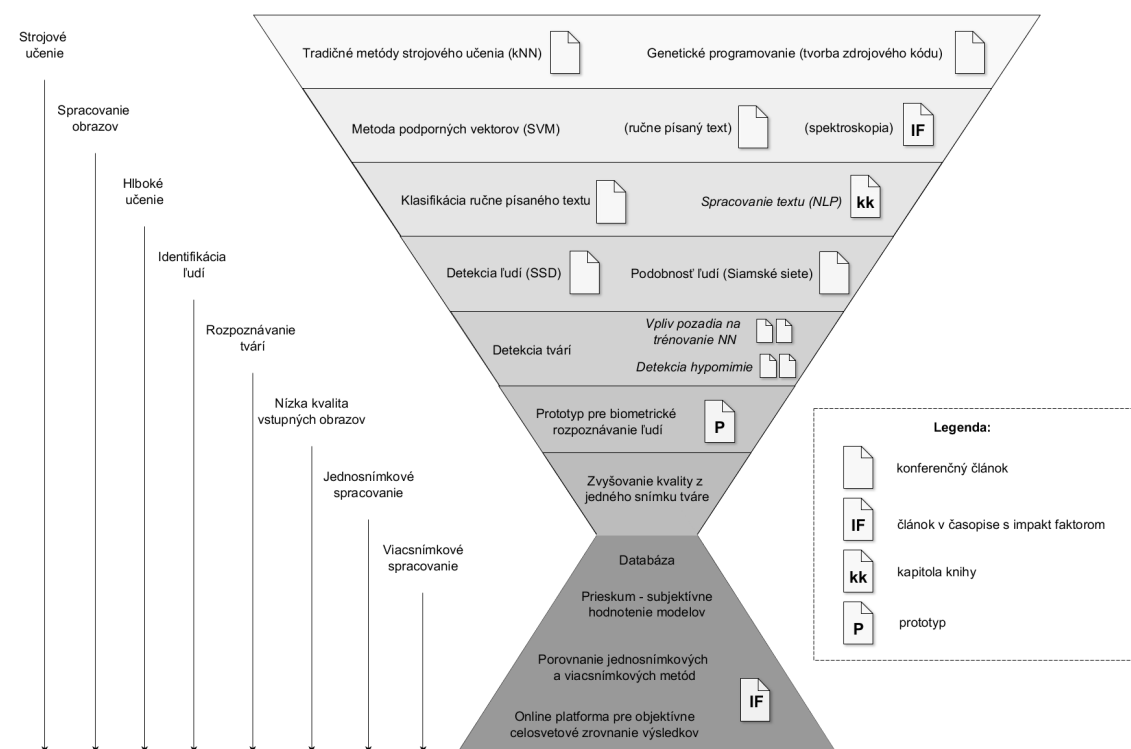
Hlavný cieľ dizertačnej práce je vytvoriť metódy založené na umelej inteligencii pre potreby biometrickej identifikácie tváre, kde súčasné metódy zlyhávajú. Jedná sa predovšetkým o prípady obrázkov s nízkou kvalitou obrazu (osvetlenie, vplyv počasia atď.) a nízkym rozlíšením. Pozornosť je upriamená na superrozlíšenie zo sekvencie snímok namiesto použitia iba jedného vstupného snímku. Tento postup je motivovaný ľudským mozgom, ktorý je schopný si vytvoriť predstavu o podobnosti človeka zo sledovania menej kvalitného videa, namiesto použitia iba jedného snímku (zastaveného) videa. Aby bolo možné dosiahnuť hlavný cieľ dizertačnej práce, je práca rozdelená na niekoľko menších, dielčích bodov:

- Zmapovať stav vedy a techniky v oblasti umelej inteligencie a strojového učenia a ich súčasné metódy (hlboké učenie), ale aj tradičné metódy. Aplikovať tradičné aj súčasné metódy na vybraný problém obrazového spracovania dát a výsledky prezentovať.
- Navrhnuť a experimentálne overiť systém biometrickej identifikácie pomocou tváre človeka pri nepriaznivých podmienkach záznamu. Systém bude zameraný na prevádzku v reálnom čase.
- Optimalizácia presnosti identifikácie tváre. Použitá metóda bude používať zvyšovanie kvality obrazu z jedného snímku osoby.
- Vytvoriť vhodný trénovací dataset pre superrozlíšenie zo sekvencie snímok obsahujúcej snímky tvárí. Požiadavky na dataset sú: objemová zrovnateľnosť s prednými svetovými datasetmi, snímky z rôznych snímacích zariadení a prostredí, rôzne nasvetlenie, vzdialenosť a uhol, osoby rôzneho veku, etnika.
- S pomocou vytvoreného datasetu experimentálne porovnať viacsnímkové a jednosnímkové metódy. Dosiahnuté výsledky prezentovať a zhodnotiť pomocou vhodných metód na určovanie podobnosti obrazov. Vytvoriť online platformu pre objektívne celosvetové zrovnanie výsledkov superrozlíšenia rôznych výskumných tímov.

Práca má za cieľ spraviť prínos do oblasti superrozlíšenia tváre zo sekvencie niekoľkých menej kvalitných snímok. Táto oblasť je málo preskúmaná a neexistuje vhodná trénovacia množina, ktorá by slúžila ako základ pre experimenty tímov z celého sveta. Motiváciou výsledného systému je pomoc pri policajnom pátraní po páchateloch závažných trestných činov (lúpežné prepadnutia, napadnutia atp.).

2 Navrhnuté metódy riešenia

Výskumná činnosť bola od začiatku zameraná na umelú inteligenciu a obrazové spracovanie dát. Všetky jednotlivé výskumné oblasti sa navzájom prelínajú a stavajú na spoločných základoch. Táto práca sa zaoberá problémom rozpoznávania ľudí z nekvalitných záznamov a hľadá spôsob, akým vylepšiť súčasný stav poznania v tejto oblasti. Práca postupne prechádza návrhom riešenia od všeobecnejších problémov ku konkrétnemu problému – prínosu v oblasti rozpoznávania osôb z nekvalitných snímok za pomoci superrozlíšenia zo sekvencie snímok tváří. Pred vymedzením sa na túto konkrétnu oblasť však bolo nutné vykonať ďalšie potrebné experimenty, ktoré stavajú na rovnakých základoch. Tieto experimenty, ktoré priniesli potrebný prehľad a zručnosti v danom odvetví, sú taktiež postupne popísané v tejto práci spoločne s dosiahnutými výsledkami. Tieto práce sa týkajú obrazového spracovania dát a začínajú základom do strojového učenia pomocou tradičných metód, neskôr prechádzajú na problematiku hlbokého učenia (klasifikácia, detekcia a podobnosť). Ďalej je vytvorený celý systém pre biometrickú identifikáciu, ktorý je následne optimalizovaný pomocou zvyšovania kvality tváre z jedného snímku osoby. Nakoniec sa práca dostane k jej hlavnému cieľu a prínosu, ktorým je superrozlíšenie tváre z niekoľkých snímok. Tento postup je spolu s dosiahnutými vedeckými výsledkami a oblasťami graficky znázornený na Obr. 2.1.



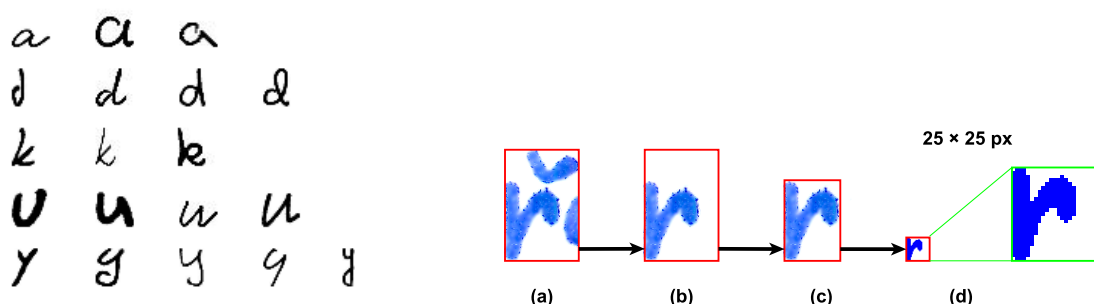
Obr. 2.1: Grafický popis riešenia od obecného ku konkrétnemu – hlavnému cieľu.

2.1 Spracovanie obrazov pomocou tradičných metód

Tradičné metódy strojového učenia môžu byť efektívne v prípade nedostatočného množstva tréovacích dát a nemožnosti použitia hlbokého učenia pre problém spracovania obrazov. Súčasťou riešenia takmer každého problému je aj tzv. predspracovanie dát. V rámci plnenia cieľov bol jeden z prvých výskumov orientovaný na rozpoznávanie ručne písaného textu, práve za použitia aj tradičných metód.

2.1.1 Rozpoznávanie ručne písaného textu

Rozpoznávanie ručne písaného textu je náročnejšie ako z tlačných dokumentov, ktoré majú definovaný štýl, veľkosť, medzery... Rozpoznávané písmo je Comenia skript [1], moderný skript písma uvedený na základné školy v Českej Republike. Ukážka Comenia skriptu a niekoľko rôznych zápisov znakov, čo komplikuje rozpoznávanie, je na Obr. 2.2 vľavo. Tieto znaky už majú odstránený šum a sú prevedené do čierneho-bieleho formátu – ide o pomerne zložité predspracovanie dát, ktoré je detailne popísané v práci [2]. Používa sa externý nástroj Tesseract [3] na detekciu znakov, nasleduje vyrovnanie náklonu znakov a ich čistenie od šumu (diakritika, časti susedných znakov atď.) a normalizovanie veľkosti, viď Obr. 2.2 vpravo.



Obr. 2.2: Ukážka variánt zápisu Comenia skriptu a predspracovania znaku [2, 4].

Po predspracovaní dát bol vytvorený dataset o veľkosti cca 12000 vzoriek obrázkov jednotlivých znakov s veľkosťou 25×25 pixelov a následne boli prekonvertované do 1-rozmerného vektoru. Vzhľadom k malému množstvu dát bol zvolený algoritmus SVM, ktorý spomedzi tradičných metód vykazuje najlepšie výsledky [5]. V práci boli vyskúšané rôzne variácie tréovania SVM modelu ako: použitie rôznych SVM jadier; vytvorenie osobitnej triedy pre tzv. „porušené“ znaky (prekrývajúce sa časti znakov, príp. diakritika); rozdelenie znakov s viacerými spôsobmi zápisu do jednotlivých tried alebo ich zlúčenie v jednej triede; vyskúšanie združovacej metódy (max-pooling) a rôznych veľkostí okien pri tréovaní. Rôzne problémy a ukážka komplexnosti rozpoznávania znakov sú zobrazené na Obr. 2.3 [4].

2.1.2 Rozpoznávanie materiálu zo spektroskopie

Ďalšia aplikácia tradičných metód bola v oblasti klasifikácie hornín za pomoci laserovej spektroskopie (LIBS). Oproti tradičným prístupom v tejto oblasti sa namiesto konverzie širokopásmového „echellogramu“ do 1D spektra využíva práve celý 2D obraz. V navrhnutom riešení, popísanom v [6], sa používa určitá oblasť záujmu (okno) echellogramov, ktorá je následne normalizovaná pomocou ekvalizácie histogramu [7]. V ďalšom kroku je vybraných 225 pixelov rôzne distribuovaných po celom obrázku (experimentovalo sa aj s 86 a 21 pixelmi), ktoré obsahujú najväčšie množstvo informácie v echellograme. Tieto pixely boli lokalizované výpočtom mediánového obrázku pre každú triedu, ktoré boli následne porovnané pre získanie rozdielového obrázku, z ktorého sa vypočítal obrázok s maximálnymi hodnotami pixelov. Tento obrázok obsahuje oblasti s vysokou intenzitou pixelov a z ich stredov sa zobrali súradnice pixelov, ktoré tvoria príznaky daného vzorku echellogramu [6].

225 bodov echellogramu značne redukuje dimenzionalitu dát – 1024×1024 (cca 1 mil. hodnôt) je reprezentovaných pomocou 255-tich hodnôt. Vytvorené datasety obsahujú 5000 vzoriek – echellogramov, ktoré sú rozdelené do trénovacej a testovacej množiny v pomere 70:30, a taktiež podľa druhu značiek – 50 druhov vzoriek (chemických prvkov) alebo 13 druhov hornín. Pomocou vytvorených dát, ale aj dát z 1D spektra, boli následne natrénované modely tradičného strojového učenia, konkrétne SVM a ako kontrast k nemu jednoduchá L2R LR [8] (Level 2 Regularized Logistic Regression), a ich presnosti boli porovnané [6].

2.2 Podobnosť obrazov s použitím hlbokého učenia

Klasifikácia

Aby bolo možné použiť hlboké učenie, resp. konvolučné neuronové siete (CNN) pre rozpoznávanie ručne písaného textu, je zvyčajne potrebné mať veľký objem trénovacích dát, čo predstavuje v tomto prípade problém. Prvou možnosťou ako rozšíriť množinu trénovacích dát je zabudovaná funkcia `Keras` [9] – `ImageDataGenerator()`, ktorá umožňuje základné transformácie obrazu, avšak na prípad písma nebolo možné použiť všetky úpravy a výsledná množina nebola dostačujúca na dosiahnutie uspokojivej presnosti. Druhou možnosťou je vytvorenie syntetických dát, dôležité však je, aby testovanie prebiehalo na reálnych dátach. Na rozšírenie datasetu boli použité ako obrazové transformácie, tak aj generovanie znakov predovšetkým pomocou horizontálnych a vertikálnych preklopení a rotácií. Zvyčajne tieto transformácie fungujú pre dvojicu obrázkov napr. horizontálne preklopenie („b“ a „d“)... Tvorili sa aj nové obrázky znakov pomocou náhodne generovaných čiar s rôznou pozíciou, šírkou, dĺžkou, náklonom a ich kompozíciou na vytvorenie znaku (napr. „T“ a „E“).



Obr. 2.3: Vyznačenie rôznych problémov pre rozpoznávanie Comenia skriptu [4].

Väčšie množstvo dát umožnilo natrénovať jednoduchú architektúru konvolučnej neurónovej siete pomocou optimalizačnej metódy SGD [10] a s použitím metódy Early stopping [11] s maximálnym čakaním 35, aby sa zabránilo pretrénovaniu, pričom tréning modelu bolo ukončené po 178-mich epochách. Následne boli na rovnakej tréningovej množine natrénované aj niektoré tradičné klasifikačné algoritmy strojového učenia a ich presnosti boli porovnané s hlbokým učením [2].

Detekcia postáv

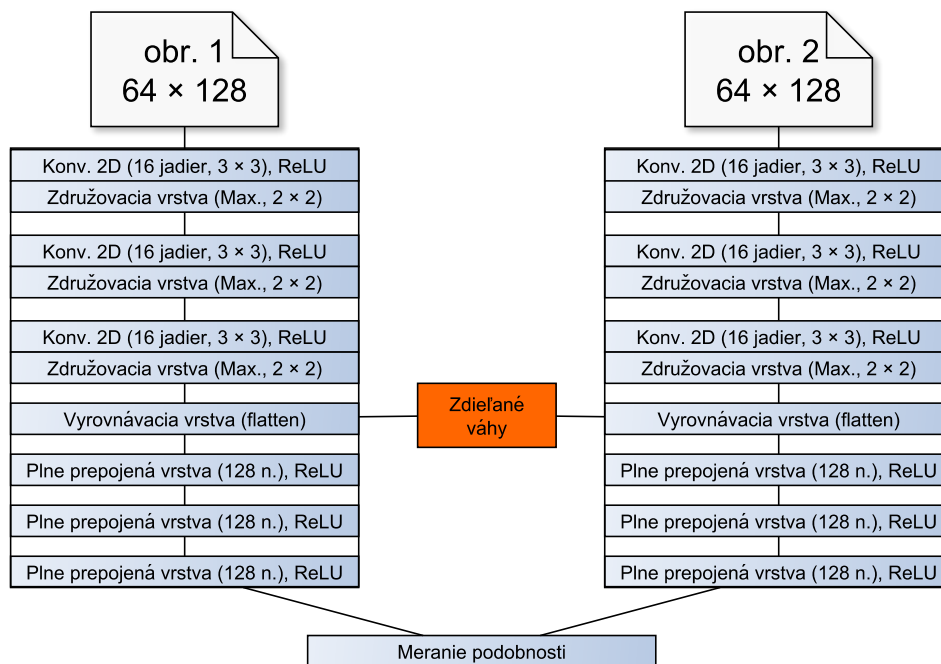
Práca sa zameriava na biometrickú identifikáciu ľudí pri nepriaznivých podmienkach záznamu, tzn. použitie v reálnom prostredí, oproti často kontrolovaným, ideálnym podmienkam používaným vo vedeckom výskume. Prvou myšlienkou bolo detekovať najskôr postavy, ktoré sú pomerovo väčšie oproti tváram daných postáv, preto by mala byť detekcia jednoduchšia (následne extrahovať tvár). Bolo vybraných niekoľko typových metód pre detekciu objektov, ktorých výkonnosť bola navzájom porovnaná a ako sa predpokladalo, hlboké učenie bolo výrazne výkonnejšie ako metódy založené na tradičných prístupoch. Preto sa ďalej použila sieť SSD [12], ktorú bolo nutné pretrénovať.

Preto bolo vytvorených cca 730 vzoriek – obrázkov označených postáv, rovnomerne rozdelených na tréningovú a validačnú množinu, a ďalších 64 vzoriek pre testovanie. Ďalej bola využitá databáza zo súťaže VOC 2007 [13], obsahujúca cca 5000 tréningových a 5000 testovacích snímokov s 20-timi triedami. V kategórii „človek“ je asi 2000 vzoriek. Kombináciou novovytvorenej databázy postáv, primárne z rozmazaných, nekvalitných snímokov a databázy VOC 2007 bolo natrénovaných niekoľko modelov, podľa typu použitých dát: boli zmazané všetky triedy okrem „človeka“ a ponechaná iba VOC databáza; boli zmazané všetky triedy okrem „človeka“ a ponechaná VOC databáza + nové vzorky; VOC databáza so všetkými triedami + nové vzorky; iba nové vzorky.

Pre tréning modelov bola použitá metóda prenosového učenia [14] (angl. transfer learning), pričom boli použité váhy modelu SSD natrénovanom na VOC 2007 data-sete. Parametre siete použité pri tréningu boli nasledovné: rýchlosť učenia 0,001, decay 0,0005, veľkosť dávky 32 a optimizér ADAM [15]. Maximálny počet tréningových epoch bol nastavený na 42000, po ktorom už sa presnosť modelu nezvyšovala.

Podobnosť ľudí pomocou Siamských neurónových sietí

V náväznosti na model detekujúci postavy bol vykonaný experiment na rozpoznávanie ľudí na základe ich celých postáv. To znamená, že takáto identifikácia ľudí by mohla pomôcť k okamžitému pátraniu po určitej osobe na základe jej oblečenia. V tomto prípade sa nehľadá asociácia k určitej triede, ale určuje sa rozdiel medzi dvoma obrazmi. Tento problém rieši tzv. Siamská neurónová sieť [16] obsahujúca dve identické neuronové siete so zdieľanými váhami, tzn. extrahujú rovnaké príznaky z obrázkov. Vo výstupnej vrstve sa používa meranie podobnosti vektorov (extrahovaných príznakov), v tomto prípade Euklidovská vzdialenosť. Použitá architektúra je zobrazená na Obr. 2.4.



Obr. 2.4: Navrhnutá architektúra Siamskej neurónovej siete [17].

Trénovací dataset bol vytvorený z videa použitého na súťaži AVSS 2007 [18] a obsahuje 56 osôb a 1179 obrázkov, rozdelených do množín na tréning (trénovacia/validačná množina) a na testovanie. Nakoľko má Siamská sieť na vstupe dva obrázky, musí byť vstup vo forme [obrázok1, obrázok2, značka]. Pozitívne vzorky boli vytvorené ako všetky kombinácie obrázkov jednotlivých postáv. Ďalej sa náhodne vytvára rovnaký počet negatívnych vzoriek, aby bol dataset vybalancovaný.

Dosiahnuté výsledky netrénovaného modelu neboli natoľko uspokojivé, aby mohol byť reálne použitý. Navyše by v uvažovanom systéme pre biometrickú identifikáciu prinášal ďalšiu réžiu v podobe nasledovnej detekcie tváre. Ďalší postup bol preto zameraný na priamu detekciu tváre, ktorá už nevyžaduje dodatočné algoritmy detekcie, čo poskytuje veľkú výhodu pri spracovaní v reálnom čase.

Detekcia tváří

Pre natrénovanie kvalitného modelu, schopného detekovať aj tváre z nepriaznivých podmienok, je nutné vytvoriť vhodný trénovací dataset s veľkým počtom snímkov s tvármi. Na to boli využité nasledujúce databázy: FDDB [19] (5171 tváří), Color-Fashion Dataset [20] (2682 postáv), BioID [21] (1521 čiernobielych tváří), Psychological Image Collection at Strling [22] konkrétne časti „2D face sets“ Aberdeen (684 snímkov), Iranian women (369 snímkov), Nottingham scans (312 snímkov) a Stirling faces (312 snímkov). Kombináciou týchto databáz vznikol nový dataset tváří pozostávajúci z 5701 vzoriek snímkov obsahujúcich viaceré tváre, rozdelený na tri množiny (trénovacia 2710 snímkov, validačná 2580 snímkov a testovacia 411 snímkov). Ukážka označeného vzorku datasetu je na Obr. 3.1 v časti 3.2.

Autori metódy YOLOv3 [23], ktorá je momentálne jednou z najúspešnejších metód pre detekciu objektov, sprístupnili natrénovaný model pre detekciu tváří verejnosti. Tento model je však naučený detekovať aj časti tváří, profilové alebo ešte viac natočené tváre, čo je pre biometrickú identifikáciu osôb nevhodné (nutnosť ďalšieho modelu/filtru). Počas trénovania siete YOLOv3, resp. vytvorenia výsledného funkčného modelu však nastávali problémy so správnou veľkosťou tzv. „anchor boxes“. Vzhľadom na veľkú variabilitu vstupných dát a napriek niekoľkým pokusom o správne nastavenie týchto hodnôt sa nepodarilo natrénovať model tak, aby výsledná detekcia správne orámovala tvár. Pre natrénovanie detekčného modelu bola preto použitá architektúra SSD [12] s predtrénovanými váhami zo siete VGG-16 [24]. Natrénovaný model nevykazoval nedostatky ako natrénovaný model YOLOv3.

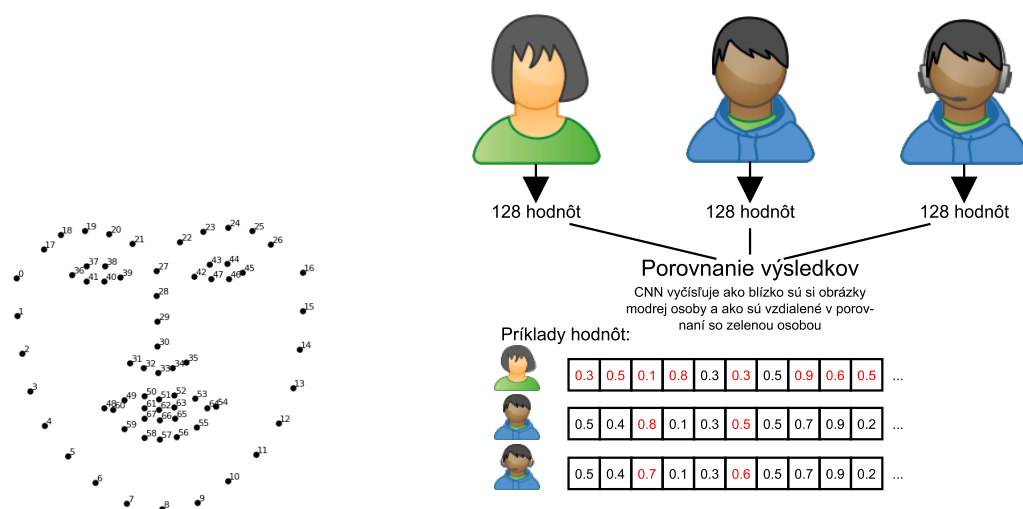
2.2.1 Systém pre biometrickú identifikáciu ľudí

Cielom tejto časti práce je navrhnúť a vyvinúť inteligentný monitorovací systém, ktorého účelom je možnosť doplniť stávajúce kamerové systémy o schopnosť automaticky upozorniť na hľadané, či podozrivé osoby, a to v reálnom čase. Motiváciou vyvíjanej technológie je reakcia na bezpečnostné riziká vo svete, hlavne v súvislosti s migračnou krízou a teroristickými útokmi. Cielom systému je poskytnúť ekonomicky výhodné riešenie, ktoré s minimálnym zásahom do súkromia a slobôd občanov bude upozorňovať na bezpečnostné riziká.

Nakoľko je celý systém zložitý, skladá sa z niekoľkých jednotlivých častí (služieb), ktoré síce spracovávajú nezávislé časti, tieto časti ale musia byť vykonané sekvenčne. Princíp celého systému je nasledujúci: na základe vstupného snímku sa detekuje tvár pomocou natrénovaného SSD modelu. Následne sú vyrezané jednotlivé časti obrázku obsahujúce tváre. Tieto tváre sú zakódované jednosmernou funkciou do vektoru o dĺžke 128 hodnôt (extrakcia príznakov) pre rozpoznávanie. Vytvorené vektory príznakov sú pomocou Euklidovskej vzdialenosti následne porovnané s vektormi

v databázi hľadaných osôb, ktorá je súčasťou systému. Kvôli rýchlosti je načítaná databáza osôb v RAM pamäti a zálohovaná na disku (obsahuje iba anonymizované dáta). Systém je koncipovaný ako serverovská časť, ktorá komunikuje cez definované rozhranie a pracuje na princípe spracovania jedného obrazu, pričom návrh a implementácia je pripravená tak, aby systém stíhal spracovávať snímky rýchlosťou videa.

Komponenta systému pre rozpoznávanie tváří používa nástroj **Face-recognition** [25], založený na hlbokoj neurónovej sieti pre rozpoznávanie ľudí, nazvanej Facenet [26], ktorá dosahuje presnosť 99,63 % na datasete LFW [27]. Moderné metódy počítajú s problémom natočenia tváří a riešia ho pomocou rôznych obrazových transformácií, aby sa zlepšilo vycentrovanie tváre [28]. Používa sa k tomu metóda na odhad významných bodov tváre [29] nachádzajúca 68 bodov, ktoré obsahuje každá tvár, viď Obr. 2.5 vľavo. Kľúčovou časťou merania podobnosti osôb je proces vytvorenia vektora príznačkov tváre (extrakcia charakteristických črtov tváre). K tomu je použitá spomínaná sieť Facenet [26], trénovaná pre generovanie 128-ich hodnôt tváre (na základe meraní autorov, 128 vykazuje najlepšie výsledky). Počas tréningu má sieť na vstupe dva obrázky rovnakej osoby a jeden obrázok inej osoby a snaží sa nájsť také hodnoty, ktoré majú minimálny rozdiel pre rovnakú osobu, a naopak čo najväčší rozdiel pre rozdielne osoby, viď Obr. 2.5 [28, 30].

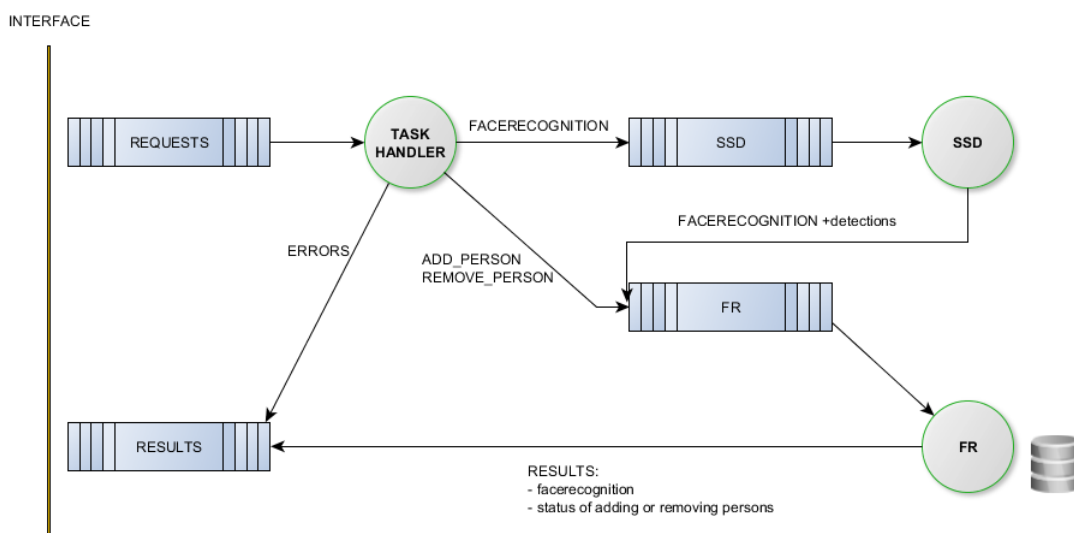


Obr. 2.5: Vizualizácia odhadu významných bodov tváre a procesu tréningu rozpoznávania tváre pomocou nástroja **Face-recognition** [28, 29, 30].

Nakoľko do služby rozpoznávania tváre prichádzajú už výrezy tváří zo služby detekcie, je nutná normalizácia vstupných detekcií. Z dôvodu výpočetnej náročnosti bolo rozhodnuté o použití metódy HOG [31] nástroja **Face-recognition**, ktorý nepotrebuje hardwarovú akceleráciu. Rýchlosť spracovania pomocou jednotlivých možností spolu s navrhnutým prístupom je zobrazená v Tab. 3.5 v časti 3.3.

Jednou z podmienok pre vytvorenie systému pre biometrickú identifikáciu je aj jeho možnosť behu v reálnom čase, čo nie je triviálna úloha, nakoľko je potrebných niekoľko úkonov od predspracovania obrázku, detekcie tváří, orezania tváří, výpočet vektorov príznačkov, vyhľadanie ich prípadnej zhody v databáze a zaslanie odpovede. Preto bol zvolený návrh pomocou architektúry tzv. mikroslužieb a frontovacieho systému, aby bol výsledný systém univerzálny, prehľadný, udržateľný, ľahký na odladovanie a vedel pracovať s odlišnými technológiami.

Hlavná časť systému obsahuje tri služby. Už spomínané jednotlivé služby pre detekciu tváří - ďalej **SSD**, služba pre rozpoznávanie tváří (normalizácia, výpočet vektorov príznačkov a vyhľadanie v databáze) – ďalej **FR** a pribudla služba na obsluhovanie prichádzajúcich požiadaviek (angl. **task handler**). Táto služba je zodpovedná za kontrolu správnosti prichádzajúcich požiadaviek, riešenie prípadných chýb a smerovanie požiadaviek. Vizualizácia architektúry celého navrhnutého systému je zobrazená na Obr. 2.6.



Obr. 2.6: Vizualizácia systému pre biometrické rozpoznávanie ľudí. Prevzaté z [32].

Ako bolo spomínané systém komunikuje cez definované rozhranie, prijíma požiadavky, ktoré postupujú do jednotlivých front a následne sú spracované pomocou jednotlivých služieb. Obrázky sú zakódované do formátu **Base64** kvôli prenosu medzi službami. Fronty fungujú na princípe FIFO (First In First Out) a v prípade pomalého spracovania, zostávajú uložené a čakajú na spracovanie vo fronte. V prípade, že to nastane, je možné pripojiť ďalšiu inštanciu služby, ktorá sa pripojí do fronty a pomáha so spracovaním. Tento spôsob je nazvaný ako škálovateľnosť a je využitý pre beh v reálnom čase.

Detekcia príznakov pre Parkinsonovu chorobu

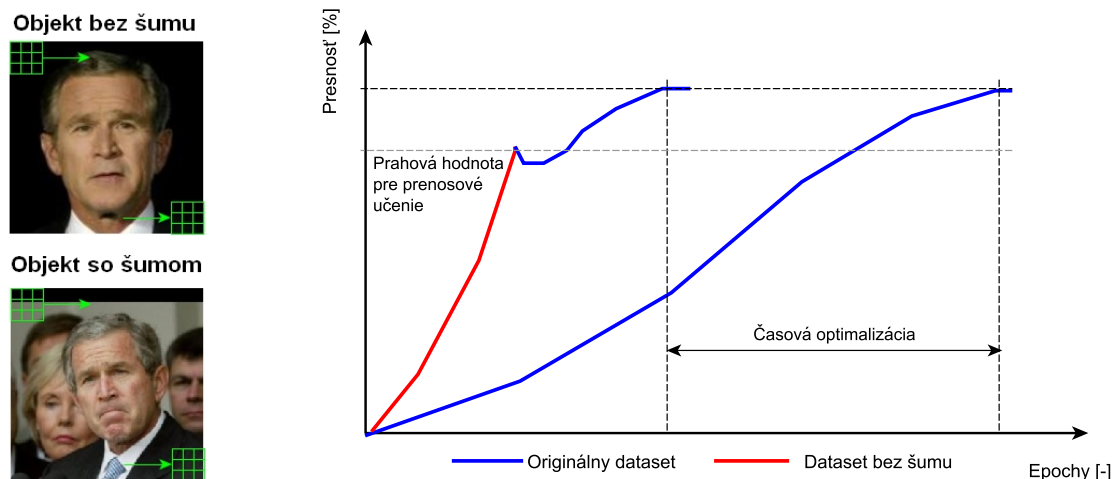
Popri študovaní problému rozpoznávania tváří prišla myšlienka využiť tieto metódy pre detekciu neurodegeneratívnych chorôb za pomoci statického snímku tváre človeka, pričom trendy v tejto oblasti sú spracovanie videa, nie statického snímku. Konkrétne sa jedná o detekciu hypomímie, ktorá je jedným z typických symptómov Parkinsonovej choroby, prejavujúca sa predovšetkým nižšou schopnosťou hýbaním svalov tváre, resp. neschopnosťou vyjadrovať emócie [33].

Úlohou experimentu bola iba jednoduchá klasifikácia bez ďalšej, detailnejšej znalosti zdravia pacientov. K dispozícii bola databáza videí s 50-timi ľuďmi trpiacimi Parkinsonovou chorobou a 50-timi zdravými ľuďmi, vybalancovaná ako vekovo, tak aj pohlavím. Aby mohol byť natrénovaný klasifikačný model, bolo potrebné vytvoriť nový, vhodnú tréningovú množinu. Z každého videa bola preto detekovaná tvár, vypočítané príznaky tváre (128 hodnôt) a priradená značka, či sa jedná o človeka zdravého alebo o človeka trpiaceho hypomímiou. Experiment sa považuje za problém binárnej klasifikácie a k tréningu modelov vzniklo 100 vzoriek, ktoré boli náhodne premiešané. Hlboké učenie nemohlo byť v tomto prípade použité vzhľadom na veľmi malý počet tréningových dát, preto boli použité niektoré z najznámejších tradičných algoritmov strojového učenia [30].

Vplyv požadania na tréning neurónovej siete

Myšlienkou pre tento experiment bola potreba rýchleho pretrénovania modelov hlbokého učenia. Výskum sa zakladá na prenosovom učení a predpokladu, že konvolučná neurónová sieť, resp. konvolučné filtre v jej jednotlivých konvolučných vrstvách, sa dokážu rýchlejšie naučiť svoje váhy, keď bude šum maximálne redukovaný oproti signálu (Obr. 2.7 vľavo). Po natrénovaní na takto upravených dátach sa s použitím prenosového učenia [14] model dotrénuje na dátach reálnych (so šumom) aby bol schopný generalizovať a nedošlo k pretrénovaniu. Hypotéza tohto experimentu je zobrazená na Obr. 2.7 vpravo. Binárna klasifikácia osôb bola vybraná pre demonštráciu tohto problému s osobami George Bush (530 obrázkov) a Colin Powell (236 obrázkov) z datasetu LFW [27]. 294 obrázkov Georga Busha bolo odstránených kvôli balancovaniu datasetu a boli vytvorené dve kópie datasetu, jedna bola ponechaná originálna a druhá s odstráneným šumom (začiernením všetkého okrem klasifikovaného objektu) [34]. Následne bolo natrénovaných niekoľko typových neurónových sietí, na ktorých sa overovala funkčnosť metódy.

Problémom pri tréningu hlbokých konvolučných sietí je množstvo dát. V prípade nedostatočného množstva tréningových vzoriek sa môže sieť ľahko pretrénovať, alebo naopak nemusí začať vôbec konvergovať (nedokáže sa učiť – nájst závislosti v dátach). Preto bola otestovaná aplikácia metódy odstránenia šumu z datasetu s malým



Obr. 2.7: Ukážka prechodu konvolučného jadra 3×3 obrázkom a hypotéza experimentu pre rýchlejšie tréňovanie s prenosovým učením [34].

množstvom vzoriek pre tréňovanie siete. Boli pripravené 4 datasety: originálne obrázky s a bez odstráneného šumu a orezané obrázky tváre s a bez odstráneného šumu. Cieľom experimentu bolo poukázať na účinnosť odstránenia šumu z obrázkov takým spôsobom, že sieť je schopná konvergovať oproti normálnemu postupu [35].

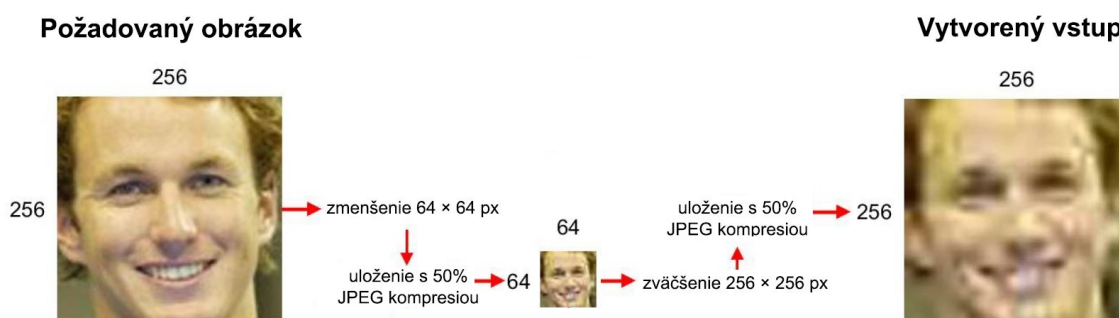
2.3 Zvyšovanie kvality snímku tváre

Kvôli optimalizácii presnosti identifikácie tvárí bola do vytvoreného systému implementovaná jednosnímková metóda, ktorá je zavolaná službou pre rozpoznávanie tváre v prípade, že nedokáže vypočítať príznaky tváre. Snímok je po vykonaní transformácie (zvýšenie kvality) následne vrátený do služby pre rozpoznávanie tváre, ktorá sa pokúsi o výpočet príznakov znovu. Aby však takáto služba (resp. model), mohla byť vytvorená, bolo potrebné vytvoriť vhodný tréňovací dataset, ktorého tvorba spočívala v niekoľkých krokoch (použitá databáza bola LFW [27]):

1. Detekcia tváre pomocou Viola-Jones detektoru [36].
2. Zmena veľkosti obrázku pomocou bikubickej interpolácie (nástroj `OpenCV`¹).
3. Z obrázku sa zobrala iba najväčšia detekcia v prípade, že obsahuje viac tvárí.
4. Vytvorilo sa viacero tréňovacích databází podľa zmenšenia na: 24×24 , 32×32 , 64×64 px a 50 % JPEG kompresiou alebo zmenšenie do náhodnej škály od 24×24 do 64×64 px a s 30–70 % JPEG kompresiou.

Vyššie popísaný postup vytvárania tréňovacieho datasetu je vizualizovaný na Obr. 2.8, konkrétne sa jedná o model 64×64 .

¹<https://opencv.org/>



Obr. 2.8: Ukážka postupu tvorenia databázy pre model 64×64 .

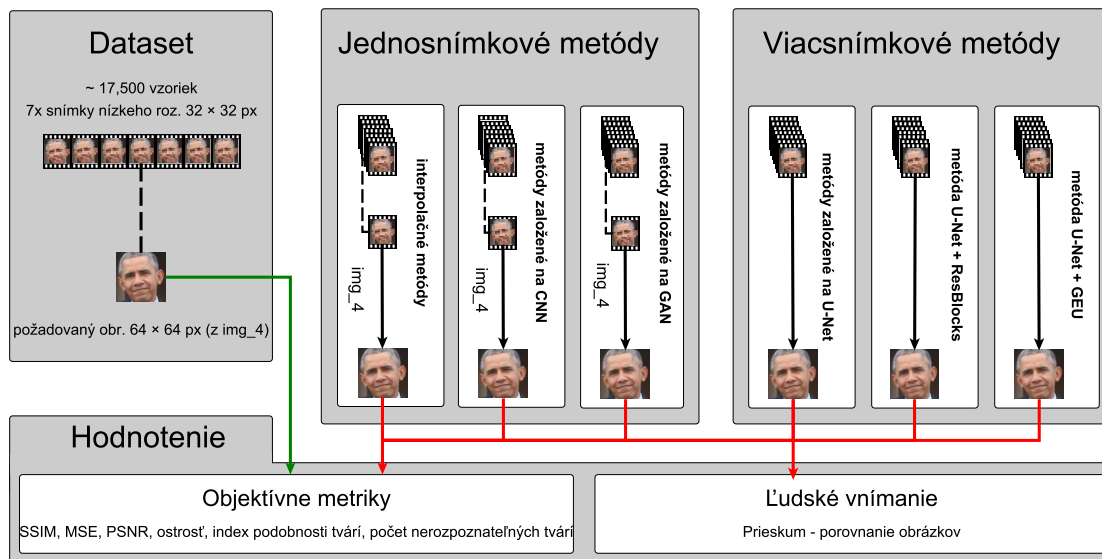
Testovacia množina bola vytvorená rovnakým spôsobom, navyše však je nutné poznamenať niekoľko bodov:

1. Zobrali sa iba tie osoby z datasetu LFW, kde boli presne dve fotky danej osoby.
2. Prvá fotka je „referenčná“ – iba detekcia tváre, orezanie a zmena veľkosti.
3. Druhá fotka je nazvaná „originál“ – prebehne detekcia a orezanie tváre, následne zmena veľkosti na 256×256 px. Ďalej zmenšenie na konkrétnu hodnotu (24, 32, 64 alebo náhodne a uloženie s 50 % alebo náhodnou kompresiou. Nasleduje zväčšenie na 256×256 px a uloženie s 50 % alebo náhodnou kompresiou.

Vytvorená trénovacia množina obsahuje 10954 snímok tváří a testovacia množina 667 snímok. Na natrénovanie modelu pre zvyšovanie kvality bola použitá neurónová sieť Pix2Pix [37], ktorá je jedna z najznámejších v oblasti transformácií typu obraz-obraz. Trénovalo sa niekoľko modelov na základe datasetu podľa rozmerov zmenšených obrázkov: 24×24 , 32×32 , 64×64 a náhodný model, pričom sa trénovalo na 100 epochách.

2.4 Zvyšovanie rozlíšenia zo sekvencie snímok

Motiváciou tejto (hlavnej) časti návrhu metód riešenia je takmer nepreskúmaná oblasť superrozlíšenia zo sekvencie snímok so zameraním sa na rekonštrukciu jedného výsledného obrázku tváre. V tejto oblasti nie sú stanovené žiadne základné štúdie, nakoľko chýbajú definované metriky hodnotenia a hlavne voľne dostupný, dostatočne veľký dataset, ktorý by mohol vedecký svet používať pre svoje experimenty a na základe ktorého by sa mohli porovnávať dosiahnuté výsledky. Nakoľko najúspešnejšie metódy superrozlíšenia sú na základe neurónových sietí, tento experiment sa nimi taktiež primárne zaoberá. Bolo natrénovaných a porovnaných 9 architektúr neurónových sietí, zahŕňajúcich metódy jednosnímkové aj viacsímkové. Spolu s 3-mi interpolačnými metódami bolo nakoniec porovnaných 12 metód. Celý experiment je graficky znázornený na Obr. 2.9 [38].



Obr. 2.9: Vizualizácia experimentu pre superrozlíšenie zo sekvencie snímkov [38].

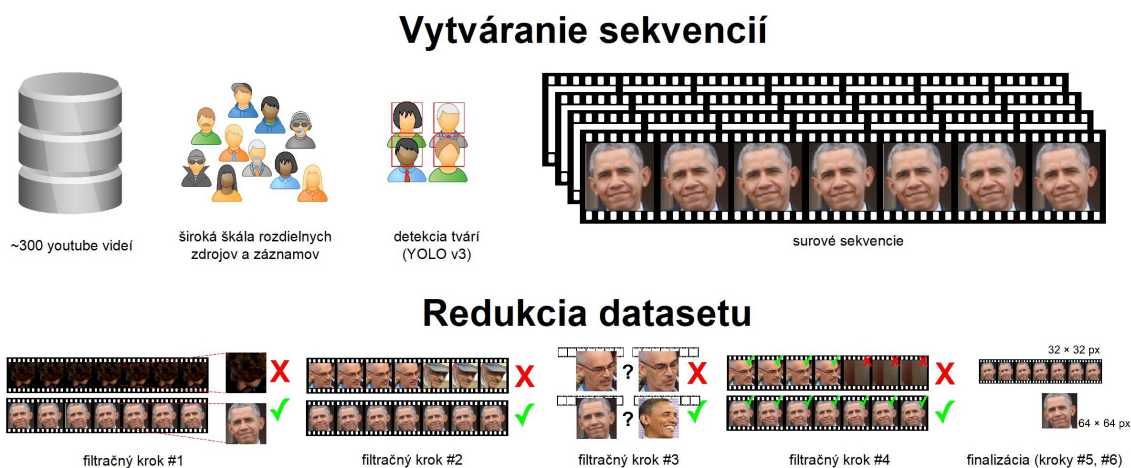
2.4.1 Návrh datasetu

V oblasti superrozlíšenia tváre zo sekvencie snímkov neexistuje dataset, pomocou ktorého by bolo možné riešiť tento problém. Známe svetové datasety obsahujú viac než 10000 vzoriek a viac než 1000 rozdielnych osôb – minimálne takéto hodnoty musel splniť vytvorený dataset. Novovytvorený dataset si kladie za cieľ vylepšiť nedostatky stávajúcich datasetov, a to hlavne pomocou záznamov z rôznych snímacích zariadení, z rôznych prostredí (nasvietenie, vzdialenosť...), záznamov ľudí rôzneho veku, oboch pohlaví, rôznych etnických skupín, nezahŕňať iba celebrity apod. a predovšetkým používať sekvencie snímkov [38].

Dataset bol vytvorený z približne 300 videí, dostupných na youtube.com. Tváre boli detekované na náhodných snímkoch videí pomocou YOLOv3 [23] a následne pomocou OpenCV sa zobrali 3 snímky pred a po. V tomto kroku bolo vytvorených 36123 sekvencií, ktoré boli redukované niekoľkými podmienkami (pomocou Face-recognition [25] nástroja založeného na sieti Facenet [26]):

1. Zo stredného obrázku musí byť možné extrahovať príznaky tváre.
2. Tváre v sekvencii musia patriť jednej osobe.
3. Odstránenie sekvencií s rovnakými osobami a podobnou scénou.
4. Všetky snímky sekvencie musia obsahovať tvár.
5. Normalizácia veľkosti sekvencie na 32×32 px a požadovaný obrázok na 64×64 px.
Náhodný výber interpolačnej metódy a hodnoty kompresie v škále (30–70).
6. Finálna kontrola – opakovanie bodu 1.

Vizualizácia tvorby datasetu je zobrazená na Obr. 2.10 korešpondujúca s popisom jednotlivých filtračných krokov.



Obr. 2.10: Vizualizácia procesu tvorby datasetu [38].

Každá jedna sekvencia obsahuje 7 snímkov osoby v nízkom rozlíšení 32×32 px, požadovaný obrázok v dvojnásobnom vysokom rozlíšení 64×64 px a informačný súbor s informáciami o zdroji videa, veľkosti JPEG kompresie a interpolačnej metóde [38].

2.4.2 Výber vhodných architektúr

Hlavným cieľom práce je superrozlíšenie tváre na základe niekoľkých snímkov (sekvencie) na rozdiel od trendov v tejto oblasti, ktorými sú superrozlíšenie z jedného snímku tváre. Tieto metódy sú považované za aktuálne najlepšie vo svete. Je to čiastočne spôsobené tým, že superrozlíšeniu zo sekvencie snímkov sa nevenuje taká pozornosť a pokrok v tejto oblasti je pomalší [39]. Porovnanie s poprednými podobnými metódami je preto problematické vzhľadom na ich veľmi malé množstvo a rozdielnosť v použití. Preto aj iné významné vedecké práce, napr. [39, 40, 41], porovnávajú výsledky s poprednými metódami superrozlíšenia z jedného snímku. Z toho dôvodu boli natrénované a porovnané jednosnímkové metódy: SRCNN [42], EDSR [43], SRGAN [44], ESRGAN [45]. Do výsledného porovnania boli tiež zahrnuté klasické, dobre známe interpolačné metódy, konkr. bikubická, bilineárna a lanczos [38].

Vytvorené viacsnímkové metódy [38] sa striktne vymedzujú proti použitiu GAN, nakoľko pre účely biometrickej identifikácie nie je možné použiť akýkoľvek generatívny prístup. Majú spoločný základ v sieti U-Net [46], ktorá je jednou z najpoužívanejších architektúr v oblasti segmentácie, avšak pre superrozlíšenia nedávala uspokojivé výsledky, preto boli vytvorené modifikácie a kombinácie s inými blokmi (ResBlock [47] a GEU blok [48]). **U-Net+ResBlock** používa oproti bežnému U-Net tzv. Sub-pixel konvolučnú vrstvu [49] namiesto Upsampling vrstvy. Pred touto sieťou je použitá $3 \times$ konvolučná vrstva s LeakyReLU a následne $5 \times$ rezidu-

álny blok. **U-Net+GEU** je podobná architektúra, akurát je namiesto reziduálneho bloku použitý GEU blok a hneď za vstupnou základnou extrakciou príznakov sú výstupy spojené a poslané do GEU blokov. Nakoľko táto architektúra produkovala tzv. „checkerboard artefakty“, bola vytvorená **U-Net+GEU2** architektúra so snahou ich potlačiť. Zmenila sa veľkosť konvolučného jadra na 7 (z pôvodných 3), aktivačná funkcia GEU je ReLU namiesto Sigmoidy a U-Net siete sa vrátila pôvodná Upsampling vrstva namiesto Sub-pixel konvolučnej vrstvy. **U-Net+GEU3** je zložená z rovnakých častí ako predchádzajúca architektúra, je však zamenené ich poradie, čo má za následok možnosť extrahovať komplexnejšie príznaky. Detailný popis architektúr je možné nájsť v [38].

2.4.3 Hodnotenie kvality

Výhodou **objektívnych metrík** je ich presná definícia, deterministický charakter, konzistentnosť a reprodukovateľnosť. Nasledovné objektívne metriky boli použité v tejto práci:

Stredná kvadratická chyba – Mean Square Error (**MSE**) [50] je definovaná ako:

$$MSE = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n (I(i, j) - \hat{I}(i, j))^2, \quad (2.1)$$

kde I je originálny obrázok a $\hat{I}$ je rekonštruovaný obrázok, obidva o šírke m a výške n pixelov.

Špičkový odstup signálu od šumu – Peak Signal-to-Noise Ratio (**PSNR**) [51]:

$$PSNR = 10 \cdot \log_{10} \left(\frac{R^2}{MSE} \right), \quad (2.2)$$

kde R je maximálna možná hodnota pixela a jednotkou PSNR je decibel [dB].

Štruktúrna podobnosť – Structural Similarity (**SSIM**) [52]:

$$SSIM(I, \hat{I}) = \frac{(2\mu_I\mu_{\hat{I}} + C_1) + (2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)}, \quad (2.3)$$

kde μ je priemer, σ^2 je variancia, σ je kovariancia signálov I , $\hat{I}$ a C_1 a C_2 sú konštanty.

Ostroť ako rozdiel kumulatívnej pravdepodobnosti rozmazania obrázkov – Cumulative Probability of Blur Detection (CPBD) [53]:

$$CPBD = P(P_{BLUR} \leq P_{JNB}) = \sum_{P_{BLUR}=0}^{P_{BLUR}=P_{JNB}} P(P_{BLUR}), \quad (2.4)$$

kde

$$P_{BLUR} = P(e_i) = 1 - \exp\left(-\left|\frac{w(e_i)}{w_{JNB}(e_i)}\right|^\beta\right), \quad (2.5)$$

kde $w(e_i)$ je šírka hrany e_i , w_{JNB} je šírka hrany zvanej JNB (Just Noticeable Blur) [54], ktorá závisí na lokálnom kontraste okolo hrany a β je parameter so stanovenou mediánovou hodnotou 3,6, určenou autormi vzhľadom k výsledkom ich práci [55].

Ďalšie metriky sú založené na open-source nástroji Face-recognition [25], ktorý využíva sieť Facenet [26]. **FR failed** udáva počet obrázkov, z ktorých nástroj nedokázal extrahovať príznaky tváre. Z obrázkov, u ktorých bolo možné extrahovať príznaky tváre testovanou metódou, sa počíta index podobnosti tvárí – **FR rate (all)**. Táto metrika počíta priemernú zhodu pomocou Euklidovskej vzdialenosti medzi extrahovaným vektorom príznakov so skutočným vektorom príznakov (predikovanej a skutočnej tváre). Aby bola táto metrika viac vypovedajúca, je definovaný index podobnosti tvárí z prieniku obrázkov, u ktorých všetky testované metódy dokázali extrahovať príznaky – **FR rate ($\cap$)** [38].

Počas experimentu však bolo zistené, že modely, ktoré **subjektívne pri ľudskom vnímaní** vykazovali relatívne dobré výsledky, pri hodnotení pomocou objektívnych metrík tomu tak nebolo. Ľudské vnímanie a hodnotenie kvality je veľmi dôležité, obzvlášť pokiaľ sa jedná o oblasť bezpečnosti a identifikácie ľudí. Z tohto dôvodu bola zavedená subjektívna metrika pomocou ľudského ohodnotenia. 126 sekvencií datasetu, vybraných tak, aby boli čo najviac rozdielne, bolo vyčlenených pre vytvorenie prieskumu, v ktorom budú ľudia vyberať najlepšiu fotku podľa ich uváženia. Prieskumu sa zúčastnilo 100 ľudí rôzneho veku, z rôznych krajín, rôzneho zamerania, napr. študenti vysokých škôl, IT odborníci, dokonca niekoľko expertov v danej oblasti, ale aj ľudia, ktorí nepracujú s informačnými technológiami. Ľudia sa prieskumu účastníci anonymne. Prieskum bol vytvorený ako webová stránka typu anketa, kde boli v jednotlivých sekciách zobrazené výsledné obrázky všetkých 12-tich metód pre jednu sekvenciu. Názvy metód samozrejme neboli k dispozícii [38].

2.4.4 Návrh testu pre verejnosť

Táto práca taktiež pojednáva o navrhnutí komplexného systému – riešenia, ktoré by prinieslo väčšiu pozornosť tejto problematike. Jedná sa predovšetkým o niekoľko častí, ktorými sú už spomínané vytvorenie a zverejnenie datasetu, ďalej vytvorenie počiatočných výsledkov na tomto datasete a nadefinovanie určitého objektívneho spôsobu hodnotenia výsledkov. Cieľom je mať všetky informácie prehľadne na jednom mieste, snažiť sa o automatizáciu hodnotenia výsledkov a možnosť zobrazovania vedúcich tímov. Z toho dôvodu bol navrhnutý MLFDB server – webová stránka², ktorá nielen poskytuje dataset, ale taktiež ponúka overenie natrénovaných modelov, aby sa zabezpečila objektivita hodnotenia modelov a zamedzilo sa prispôbovaniu výsledkov na testovací dataset (počet pokusov je obmedzený) [38].

²<http://splab.cz/mlfdb/>

3 Overenie navrhnutých metód

Výsledky navrhnutých metód riešenia, plniace ciele dizertačnej práce, sú popísané a rozdelené do jednotlivých častí v tom istom slede ako sú popísané v kapitole 2 – Navrhnuté metódy riešenia.

3.1 Tradičné metódy

3.1.1 Rozpoznávanie ručne písaného textu

Rozpoznávanie ručne písaného textu sa zameriava na konkrétny Comenia skript, ktorý sa začína používať na základných školách v Českej Republike. Bolo vyskúšaných niekoľko variánt návrhu riešenia, ako napr. trénovanie modelu s väčším počtom tried (zahŕňajúc rôzne formy zápisu) alebo vytvorenie osobitnej triedy, pomenovanej ako „porušená“, pre znaky, ktoré sú obtiažne detekovateľné (napr. prekrývajúce sa znaky, vplyv diakritiky alebo len časti znakov). Dosiahnuté výsledky klasifikátoru SVM s použitím rôznych jadier je možné vidieť v Tab. 3.1.

Tab. 3.1: Presnosť [%] algoritmu SVM s a bez triedy pre „porušené“ znaky [4].

Jadro	28 tried (znaky spolu)	53 tried (znaky oddelene)	27 tried (znaky spolu)	52 tried (znaky oddelene)
Polynomiálne	70,88	61,70	72,73	62,85
Sigmoid	79,67	76,77	89,22	83,83
RBF	90,36	85,05	92,86	86,49
Lineárne	83,21	79,76	88,86	83,64

Najlepšia dosiahnutá presnosť klasifikácie bola 92,86 % bez použitia extra triedy pre „poškodené“ znaky a bez rozdeľovania rovnakých znakov s iným zápisom do ďalších tried [4].

3.1.2 Rozpoznávanie materiálu zo spektroskopie

Výsledné modely klasifikujú buď konkrétne horniny (13 tried), alebo jednotlivé chemické prvky (50 vzoriek). Klasifikácia prebiehala pomocou natrénovaných modelov lineárneho L2R LR a nelineárneho SVM algoritmu. Tieto modely boli taktiež natrénované aj pomocou tradičného prístupu v tejto oblasti, ktorým je prekonvertovanie echellogramu do 1D spektra. Dosiahnuté výsledky navrhnutého riešenia preukázali vysokú výkonnosť a sú porovnané v Tab. 3.2, pričom najlepšie výsledky sú zvýraznené [6].

Tab. 3.2: Presnosť [%] SVM a L2R LR klasifikácie materiálov [6].

Model	Dimenzia dát	Počet tried: 50 vzoriek		Počet tried: 13 hornín	
		SVM	L2R LR	SVM	L2R LR
1D spektrum	35000×5000	96,5	95,5	98,9	98,9
225 pixelov	225×5000	98,13	97,2	99,2	99,3
86 pixelov	86×5000	97,7	97,3	99,13	98,9
21 pixelov	21×5000	96,5	92,5	98,5	95,4

Najlepšie výsledky boli dosiahnuté pomocou SVM pri klasifikácii do 13-tich tried a s 225-timi vybranými pixelmi [6].

3.2 Hlboké učenie

Klasifikácia ručne písaného textu

Pomocou navrhnutých metód navyšovania dát vznikol nový, rozšírený dataset, pomocou ktorého už bolo možné trénovanie modelov hlbokého učenia (obsahujúci 805–1085 vzoriek pre každú kategóriu [56, 57]). Vzhľadom na to, že niektoré znaky boli umelo generované, v testovacej množine boli ponechané iba originálne znaky pre správne zhodnotenie výkonnosti modelov [2]. Natrénovaných bolo niekoľko metód strojového učenia s hľadaním parametrov pomocou metódy tzv. „grid search“ a najlepšie modely boli vybrané a porovnané v Tab. 3.3.

Tab. 3.3: Porovnanie presnosti metód strojového učenia [2].

Metóda	Lin. klasifikátor	<i>K</i> -NN	Rozh. strom	Náhodný les	SVM	CNN
Presnosť [%]	61,06	82,59	51,84	69,57	81,03	90,04

Najlepšia dosiahnutá presnosť – 90,04 % bola pomocou hlbokého učenia oproti najlepšej presnosti 82,59 % tradičných metód (*k*-nn), čo potvrdilo, že hlboké učenie dokáže lepšie spracovať obrazové informácie ako tradičné metódy strojového učenia. Taktiež bolo dokázané, že je možné natrénovať model na umelo vytvorených dátach a pri testovaní na reálnych dátach funguje s obstojnou presnosťou [2].

Aplikácia hlbokého učenia v spracovaní prirodzeného jazyka – nakoľko sa práca venovala problému rozpoznávaniu písma, ktoré spadá do oblasti spracovania prirodzeného jazyka, bol aj tejto problematike venovaný priestor nad rámec cieľov práce. Výstupy tohto výskumu sú spracované pod názvom „A Novel Big Data Approach for Text Supported Service Operations Management“ ako kapitola knihy „Big Data for Service Operations Management“ od vydavateľa Springer, ktorá ku dňu písania dizertačnej práce čaká na publikovanie (prijaté 23.11. 2020).

Detekcia postáv

Metóda pre detekciu ľudí bola prvým krokom pri návrhu systému pre biometrickú identifikáciu ľudí za nepriaznivých podmienok záznamu. Pre tento účel bolo natrénovaných niekoľko modelov na základe vytvorených datasetov pre nepriaznivé prostredie. Celkové výsledky presnosti a falošných detekcií natrénovaných modelov v závislosti na detekčnom prahu, sú porovnané v Tab. 3.4.

Tab. 3.4: Porovnanie presnosti [%] a falošne pozitívnych detekcií postáv [%] [58].

Detektor	Prah 0,5		Prah 0,4		Prah 0,3		Prah 0,2		Prah 0,1	
	Acc.	Fal. d.	Acc.	Fal. d.	Acc.	Fal. d.	Acc.	Fal. d.	Acc.	Fal. d.
HOG	34,71	15,10	—	—	—	—	—	—	—	—
Viola-Jones	17,80	10,43	—	—	—	—	—	—	—	—
C4	28,58	13,31	—	—	—	—	—	—	—	—
SSD + nové vzorky	47,27	5,76	53,13	6,47	59,04	10,43	69,13	23,02	85,14	70,50
SSD iba nové vzorky	53,37	11,87	60,44	18,35	66,61	20,50	75,86	34,53	89,59	80,58
SSD + nové vz. 1 kat.	43,06	2,78	52,30	8,27	59,56	9,71	67,16	17,27	81,87	70,14
SSD originál 1 kat.	43,80	3,60	50,76	8,63	60,36	12,59	69,89	36,69	90,48	86,33
SSD originál	45,30	6,83	51,32	7,91	54,75	14,39	60,91	22,30	81,85	62,95

Presnosť (**Acc.**) reprezentuje priemernú hodnotu zo všetkých individuálnych obrázkov ako pomer medzi predikovanými detekciami a označenými skutočnými detekciami. Pri znižovaní prahu detekcie stúpal počet falošných detekcií (**Fal. d.**) exponenciálne. Najlepšie výsledky vykazoval model pri použití detekčného prahu 0,3. Ďalej sa docielil 11,98 % nárast presnosti oproti originálnemu SSD modelu [58].

Podobnosť ľudí pomocou Siamskej neurónovej siete

Presnosť natrénovaného modelu bola 81,66 % na validačnej množine a 70,28 % na nezávislej testovacej množine. Relatívne nízka presnosť modelu môže byť spôsobená malým množstvom tréningových dát, čo vedie pravdepodobne k pretrénovaniu modelu. Na druhú stranu nebolo možné zjednodušiť architektúru siete, pretože by model nebol schopný konvergenzie [17]. Táto metóda je však nevhodná pre biometrickú identifikáciu ľudí, pretože vôbec nezohľadňuje hlavné črty tváre.

Detekcia tváří

Vykonaný výskum jednoznačne ukázal, že detekovať priamo tváre je výhodnejšie či už z pohľadu rýchlosti systému, ktorá je kritická pre beh v reálnom čase, tak aj z pohľadu presnosti detekcie. Detekcia postáv by prinášala značnú mieru ďalšieho spracovania podobne ako spomínaný problém s modelom natrénovaným pomocou siete YOLOv3 [23], ktorý zanáša do systému veľké množstvo nadbytočného šumu, ktorému sa predišlo práve použitím siete SSD [12]. Natrénovaný model dosahuje

94,61 % presnosť na testovacej množine, ktorá vôbec nebola použitá pre tréning, tzn. model tieto obrázky nikdy nevidel. Ukážka výsledku modelu je na Obr. 3.1.



Obr. 3.1: Príklad detekcie tvárí natrénovaným modelom.

3.3 Systém pre biometrickú identifikáciu ľudí

Navrhnutý a vytvorený systém dokáže dopĺňať možnosti stávajúcich kamerových systémov o schopnosť automaticky monitorovať a analyzovať verejné priestory a je schopný s minimálnym zásahom do súkromia a slobôd občanov upozorňovať na výskyt podozrivých a hľadaných osôb na monitorovaných oblastiach. Používa dve hlboké neurónové siete, preto využíva možnosť hardwarovej akcelerácie pomocou GPU, aby dokázal bežať v reálnom čase. Pre detekciu tvárí sa používa vlastný natrénovaný model SSD [12], **face-recognition** nástroj používa metódu HOG [31] pre detekciu tvárí a sieť Facenet [26] pre rozpoznávanie tvárí. Porovnanie priemerných časov jednotlivých operácií je zobrazené v Tab. 3.5.

Tab. 3.5: Porovnanie priemerných časov operácií na CPU a GPU (Titan Xp) [32].

Operácia	CPU [s]	GPU [s]
Detekcia metódou SSD	0,76	0,02
Extrakcia príznakov tváre	0,35	0,01
HOG detekcia a extrakcia príznakov tváre	1,26	0,35
SSD detekcia a extrakcia príznakov tváre	—	0,031

Posledný záznam v tabuľke poukazuje na výkonnosť a efektívnosť hardwarovej akcelerácie hlbokého učenia na GPU. Dosiahnutý čas 0,031 sekúnd je čas potrebný

na detekciu tváre pomocou SSD modelu a následné vypočítanie príznakov tváre pomocou služby rozpoznávania tváre (FR). Výhodou systému je aj jeho škálovateľnosť, ktorá umožňuje jeho nasadenie aj do prostredí s niekoľkými kamerami. Za použitia jednej inštancie služby na detekciu tváre (SSD) je rýchlosť spracovania 40 fps, pri použití dvoch inštancií je to 60 fps. Služba na rozpoznávanie tváre (FR) pracuje rýchlosťou 13 fps, za použitia štyroch inštancií sa dosahuje rýchlosť spracovania približne 50 fps (Nvidia Titan XP) [32]. Systém pre biometrickú identifikáciu potrebuje pre svoj beh v reálnom čase špecifický hardware a kombinácia s modernými metódami strojového učenia robí z neho prototyp [59].

3.3.1 Detekcia príznakov pre Parkinsonovu chorobu

Hypomímia je jedným z hlavných príznakov Parkinsonovej choroby a cieľom tejto štúdie bolo oproti stávajúcim metódam využívajúcim video, poukázať na možné použitie len jedného statického snímku. Výsledné modely sú uvažované ako binárna klasifikácia a predpovedajú, či osoba trpí hypomímiou alebo nie. Výsledky porovnania presnosti modelov sú v Tab. 3.6.

Tab. 3.6: Presnosť modelov detekcie hypomímie [30].

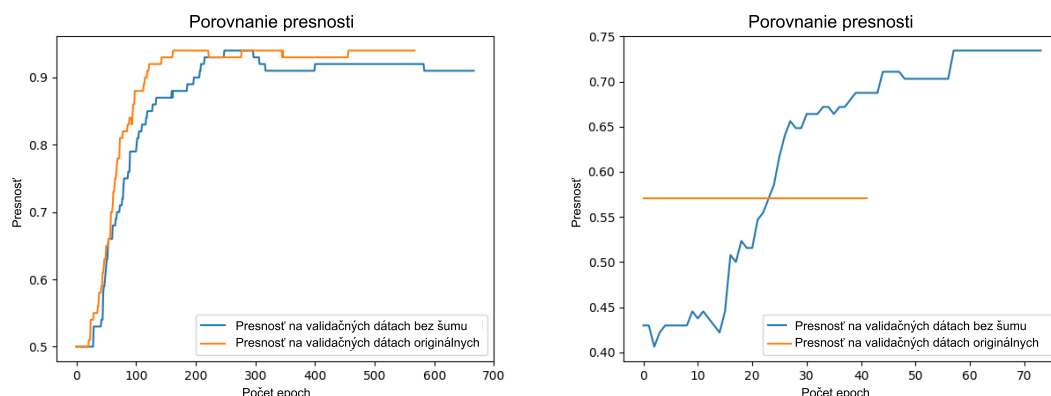
Metóda	Presnosť [%]			
	Krížová val.	Pomer 50/50	Pomer 60/40	Pomer 70/30
<i>K</i> -najbližších susedov	63,37	58,82	65,57	71,83
XGBoost	59,41	50,98	67,21	83,10
Rozhodovací strom	67,33	50,98	65,57	73,24
Náhodný les	58,42	60,70	77,05	85,92
SVM	35,64	49,02	57,38	64,79

Najlepšia presnosť bola 67,33 % a to algoritmom rozhodovací strom. V prípade overenia presnosti na testovacej množine dosahuje najlepšie výsledky algoritmus náhodný les, ale presnosť stúpa so zväčšujúcou sa tréningovou množinou – modely budú pravdepodobne pretrénované, preto sa uvažuje s najvyššou presnosťou 67,33 %. Presnosť ale nebola cieľom, tým bolo poukázanie na možnosť detekcie hypomímie zo statického obrazu [30].

3.3.2 Vplyv pozadia na tréning neurónovej siete

Tento experiment pojednáva o čase potrebnom na tréning a nájdenie spôsobu, akým uľahčiť neurónovej sieti problém, ktorý má riešiť. Navrhnutým postupom vo výsledku nastala akcelerácia učiaceho procesu v priemere o cca 69 %. Pomocou navrhnutej metódy – odstraňovania šumu, je však možné natréňovať modely, ktoré

by za použitia normálneho postupu nekonvergovali (málo dát, nevhodné tréningové dáta. . .) viď Obr. 3.2.



Obr. 3.2: Proces tréningovania a závislosť presnosti na počte epoch pri použití validačnej množiny, pričom vľavo boli použité orezané obrázky a vpravo originálne, neorezané obrázky [35].

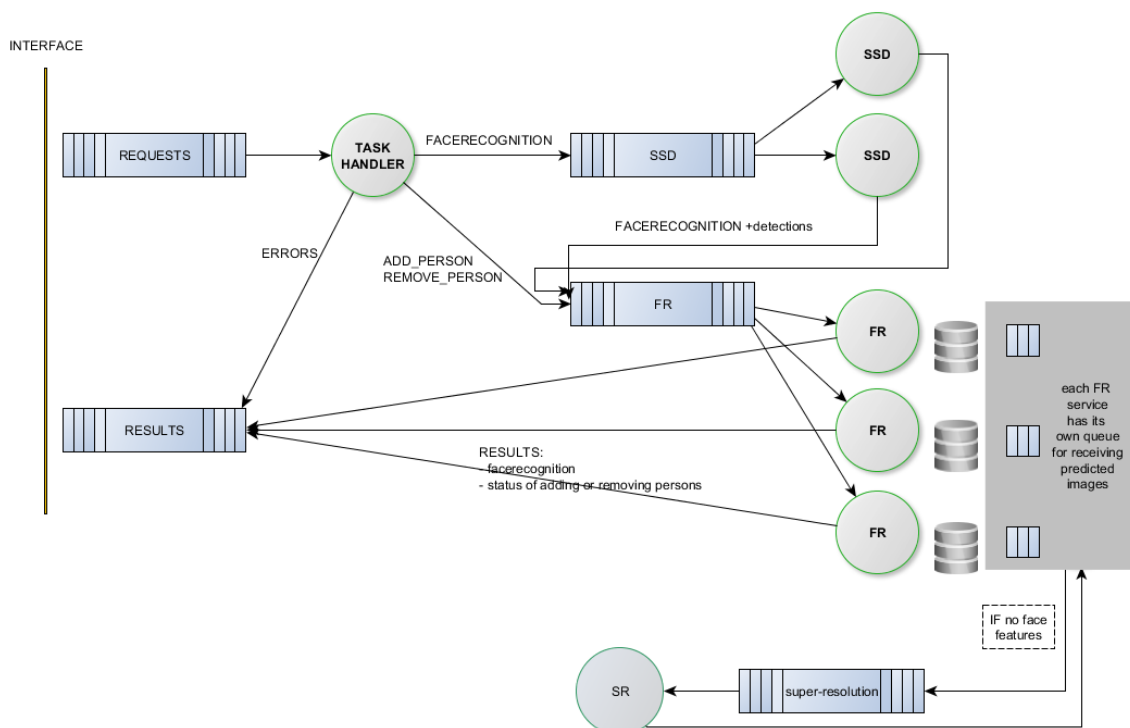
Na grafe vľavo sú zobrazené priebehy tréningovania modelov na datasete s takmer žiadnym šumom a preto sa v oboch prípadoch dokázali modely natréňovať. Na grafe vpravo sú zobrazené priebehy tréningovania na snímkoch obsahujúcich veľa šumu a sieť tu nedokázala konvergovať bežným spôsobom [34, 35].

3.4 Jednosnímkové spracovanie

3.4.1 Zvyšovanie kvality tváří

Systém pre biometrickú identifikáciu bolo nutné optimalizovať z hľadiska presnosti rozpoznávania – zvýšenie kvality tváre z obrázku, kde je obtiažna identifikácia, resp. doteraz nebola automaticky možná. Možnosť škálovateľnosti systému umožňuje jeho beh v reálnom čase a vytvorená optimalizačná služba sa používa iba v prípade, keď stávajúci systém nedokáže z obrázku tváre vypočítať príznaky. Z tohto dôvodu sa používa iba jedna inštancia tejto služby a jej implementácia do stávajúceho systému je znázornená na Obr. 3.3.

Aby takáto optimalizačná služba mohla vzniknúť, bola vytvorená vhodná tréningová množina (tréningová/testovacia množina = 10954/667) obsahujúca páry tvárí (nízka a vysoká kvalita). Na vytvorenom datasete boli natréňované 4 modely, ktoré boli porovnané aj s bikubickou interpoláciou, ktorá je jednou z najpoužívanejších metód pre zmenu veľkostí obrazov v súčasnosti, viď Tab. 3.7. Index podobnosti tváří je v tabuľke označený ako **pred. rate** pre priemer natréňovaných modelov a **bik. rate**



Obr. 3.3: Implementácia procesu zvyšovania kvality tváří do stávajúceho systému. Ohľad je braný aj na beh systému v reálnom čase a možnosť jeho škálovateľnosti.

pre priemer za použitia bikubickej interpolácie. Počet detekcií, je označený obdobne – **pred. det.** a **bik. det.** Percentuálny nárast modelov je označený ako **diff.**

Tab. 3.7: Porovnanie natrénovaných modelov pre zvyšovanie kvality tváre.

	Originálne obrázky					
Model	Pred. rate	Bik. rate	Pred. det	Bik. det.	Diff. rate [%]	Diff. det. [%]
24 × 24	0,5820	0,5394	656	495	−7,85	32,53
32 × 32	0,5096	0,4744	654	605	−7,42	8,10
64 × 64	0,2742	0,2310	651	650	−18,72	0,15
Náhodný	0,3938	0,3673	652	625	−7,21	4,32
	Referenčné obrázky					
24 × 24	0,6325	0,5933	644	486	−6,20	32,51
32 × 32	0,5885	0,5556	642	593	−5,92	8,26
64 × 64	0,4719	0,4553	635	634	−3,66	0,16
Náhodný	0,5297	0,5106	639	612	−3,73	4,41

Ako je vidieť z výslednej tabuľky, natrénované modely vykazujú lepšie výsledky čo sa týka percentuálneho rozdielu v miere podobnosti osôb v porovnaní s bikubic-kou interpoláciou. Dôležitým výstupom tohto experimentu je predovšetkým lepšia schopnosť natrénovaných modelov detekovať tváre. Pre reálne nasadenie do stáva-

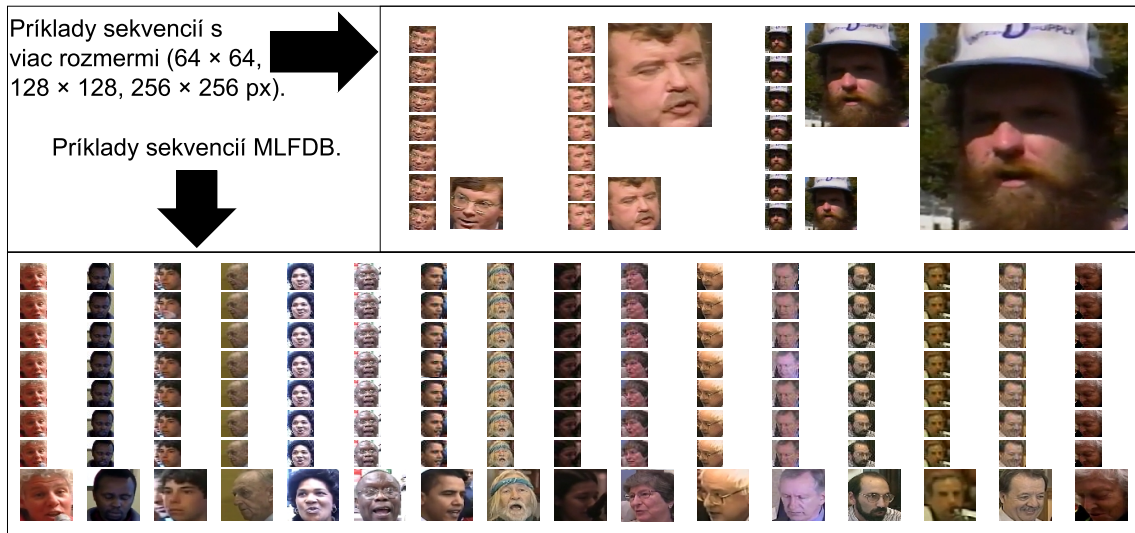
júceho systému (prototypu [59]) bol ako najspoľahlivejší model vybraný ten s trénovacou množinou označenou ako náhodná – kvôli vyššej variabilite, robustnosti, generalizácii a praktickému využitiu.

3.5 Spracovanie sekvencie snímok

Hlavný cieľ práce sa sústreďuje na oblasť superrozlíšenia zo sekvencie snímok, ktorá nie je tak výrazne preskúmaná ako oblasť superrozlíšenia za pomoci jedného snímku, obzvlášť so zameraním sa na tváre. Preto sa táto práca zameriava na riešenie, ktoré by prinieslo väčšiu pozornosť tejto problematike. Jedná sa o časti – vytvorenie a zverejnenie datasetu, vytvorenie počiatočných výsledkov, nadefinovanie objektívneho spôsobu hodnotenia modelov a umožnenie rôznym tímom z celého sveta porovnávať výsledky.

3.5.1 Dataset

Dataset sa zameriava na nekvalitné vstupné dáta (video), z ktorého je snaha o čo najlepšiu rekonštrukciu tváre do vyššieho rozlíšenia, pričom v súčasnosti neexistuje podobný dataset. Dataset obsahuje **17426** sekvencií (vzoriek), ktoré sú rozdelené pomerom 70/15/15 na trénovaciu množinu, verejnú testovaciu množinu, privátnu testovaciu množinu a zvlášť vytvorenú množinu pre prieskum subjektívneho hodnotenia. Privátna testovacia množina obsahuje iba sekvencie bez požadovaných obrázkov (labelov). Dataset obsahuje najrôznejšie sekvencie, za použitia rôznych snímacích zariadení, rôzneho nasvietenia, uhlov, vzdialenosti, kvality atď. Taktiež obsahuje záznamy z reálnych prostredí a veľké spektrum ľudí, rôzneho veku, etnických skupín, s okuliarmi alebo bez apod. Predovšetkým eliminuje nedostatky iných známych datasetov, ktoré sú vytvorené prevažne z celebrit a sú často tvorené v kontrolovanom prostredí. Každá vzorka datasetu obsahuje sekvenciu siedmich obrázkov nízkej kvality a v nízkom rozlíšení – 32×32 px a k nim odpovedajúci obrázok (label) vysokej kvality a v dvojnásobného rozlíšenia – 64×64 px. Pokiaľ to kvalita vstupného videa umožňovala, boli vytvorené aj výsledné obrázky (labels) vo vyššej kvalite – 128×128 px (12165 sekvencií) a 256×256 px (4199 sekvencií). Jedna vzorka ďalej obsahuje textový súbor s informáciami o zdroji videa, presnom čísle snímku (angl. frame), použitej interpolácii pri zmenšovaní obrázkov a veľkosti JPEG kompresie. Príklad niekoľkých vzoriek datasetu a jeho rôznorodosť je zobrazená na Obr. 3.4 [38].



Obr. 3.4: Ukážka sekvencií datasetu MLFDB [38].

Nakoľko dataset obsahuje veľké množstvo vzoriek, nebolo možné manuálne zistiť koľko rôznych osôb obsahuje. Preto bol tento počet odhadnutý na 6000–7000 pomocou automatizovanej metódy na základe nástroja **Face-recognition** [25, 38].

3.5.2 Porovnanie metód

Metódy jednosnímkového superrozlíšenia su považované za najpresnejšie v súčasnosti, preto boli niektoré z nich implementované, natréňované a porovnané s metódami viacsímkovými. Na hodnotenie výkonnosti jednotlivých metód boli použité objektívne metriky popísané v časti 2.4.3, ktorými sú: **SSIM**, **MSE**, **PSNR**, **ostrosť** počet nerozpoznatelných tvárí (**FR failed**) a indexy podobnosti tvárí (**FR rate (all)**), a prieniku obrázkov, ktoré dokázali rozpoznať všetky metódy (**FR rate ($\cap$)**). Všetkých 12 metód je porovnaných v Tab. 3.8 [38].

Tab. 3.8: Výsledky jednosnímkových a viacsímkových metód – priemer [38].

Metóda	SSIM	MSE	PSNR	Ostrosť	FR rate (all)	FR rate ($\cap$)	FR failed
bicubic	0,806	227,196	25,654	0,084	0,476	0,473	674/2500
bilinear	0,809	217,368	25,769	0,158	0,471	0,468	693/2500
lanczos	0,802	238,247	25,504	0,024	0,483	0,480	700/2500
EDSR	0,780	286,999	24,687	-0,025	0,495	0,494	701/2500
SRCNN	0,815	216,940	25,771	0,209	0,461	0,456	643/2500
SRGAN	0,760	320,981	24,234	-0,078	0,511	0,510	771/2500
ESRGAN	0,788	281,192	24,604	-0,016	0,488	0,485	470/2500
U-Net+GEU	0,760	276,459	24,383	-0,129	0,507	0,503	317/2500
U-Net+GEU+filter	0,815	234,552	25,196	0,239	0,464	0,458	394/2500
U-Net+GEU2	0,799	256,511	24,878	0,031	0,472	0,467	259/2500
U-Net+GEU3	0,799	243,433	25,135	0,004	0,471	0,466	270/2500
U-Net+ResBlock	0,812	239,114	25,141	0,130	0,468	0,463	581/2500

SRCNN dosiahla najlepšie výsledky, čo sa týka počtu najlepších – objektívnych metrík, bez subjektívneho ľudského vnímania (**sub. h.**). U-Net+GEU2 výrazne prekonala SRCNN pri ľudskom vnímaní, pričom rozdiel ich objektívnych metrík je minimálny, až na FR failed a ostrosť v prospech U-Net+GEU2 [38]. Pre lepšiu čitateľnosť a zahrnutie subjektívneho hodnotenia boli výsledky normalizované – menšia hodnota je lepšia. Tieto výsledky sú zobrazené v Tab. 3.9, kde **sum1** je finálne skóre všetkých objektívnych metrík a **sum2** zahŕňa aj subjektívne hodnotenie [38].

Tab. 3.9: Porovnanie normalizovaných hodnôt metód a finálneho skóre [38].

Metóda	SSIM	MSE	PSNR	Ostr.	FR ($\cap$)	FR fail.	Sub. h.	Sum1	Sum2
bicubic	0,164	0,099	0,076	0,340	0,315	0,811	1	1,805	2,805
bilinear	0,109	0,004	0,001	0,655	0,222	0,848	0,989	1,839	2,828
lanczos	0,236	0,205	0,174	0,085	0,444	0,861	0,999	2,005	3,004
EDSR	0,636	0,673	0,705	0,703	0,089	0,863	0,977	3,669	4,646
SRCNN	0	0	0	0,872	0	0,750	0,978	1,622	2,600
SRGAN	1	1	1	1	0,315	1	0,995	5,315	6,310
ESRGAN	0,491	0,618	0,759	0,051	0,537	0,412	0,844	2,868	3,712
U-Net+GEU	0	0,572	0,903	0,532	0,870	0,113	0,570	2,990	3,560
U-Net+GEU+filter	0	0,169	0,374	1	0,037	0,246	0,578	1,844	2,422
U-Net+GEU2	0,281	0,380	0,581	0,115	0,204	0	0	1,571	1,571
U-Net+GEU3	0,281	0,255	0,414	0	0,185	0,021	0,204	1,166	1,370
U-Net+ResBlock	0,055	0,213	0,410	0,536	0,130	0,629	0,834	1,973	2,807

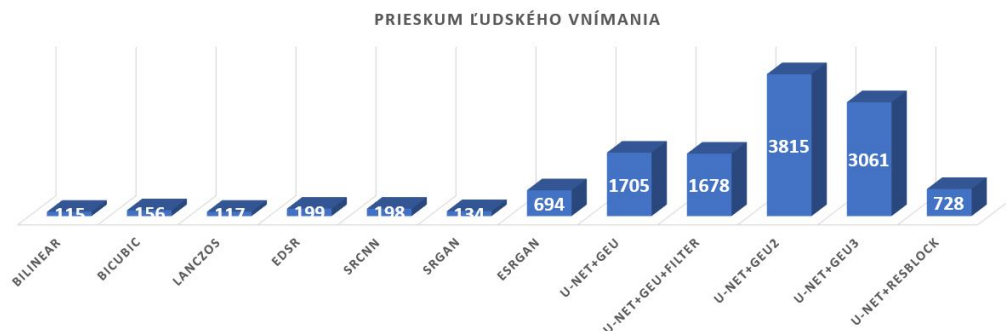
U-Net+GEU2 a U-Net+GEU3 dosahujú lepšie finálne skóre oproti metóde SRCNN, dokonca aj bez zahrnutia subjektívneho hodnotenia. Spolu so subjektívnym hodnotením sa rozdiely medzi nimi ešte prehĺbujú. Grafická ukážka dosiahnutých výsledkov jednotlivých metód je zobrazená na Obr. 3.5.



Obr. 3.5: Porovnanie metód na niekoľkých sekvenciách [38].

Všetky viacsnímkové metódy dosahujú pocitovo výrazne lepšie výsledky oproti jednosnímkovým metódam a interpolačným prístupom [38].

Subjektívne hodnotenie kvality modelov hrá významnú úlohu, nakoľko predpokladané aplikácie uvažujú použitie v oblasti bezpečnosti, preto je dôležité, ako vnímajú výsledky metód ľudia. Výsledky prieskumu sú prezentované na Obr. 3.6.



Obr. 3.6: Výsledky subjektívneho hodnotenia kvality modelov. Prieskum obsahoval 126 porovnaní 12-tich metód a účastnilo sa ho 100 ľudí [38].

Metóda U-Net+GEU2 vykazuje najlepšie výsledky v ľudskom vnímaní a nasledujúcou metódou je U-Net+GEU3. Po výraznom rozdiely je ďalšiou metódou U-Net+ResBlock s podobným výsledkom ako má ESRGAN metóda, ktorá ako jediná vyčnieva spomedzi jednosnímkových metód. Metódy s dobrými výsledkami objektívnych metrík v ľudskom vnímaní zlyhali oproti metódam, ktoré nevykazovali až tak dobré výsledky pri objektívnych metrikách [38].

3.5.3 Celosvetové porovnanie výsledkov

Na webových stránkach Signal Processing Laboratory (SPLAB) Vysokého učenia technického v Brne je dostupný online hodnotiaci systém pre metódy superrozlíšenia tváre zo sekvencie snímok. Akýkoľvek výskumný tím zo sveta si môže nechať objektívne otestovať svoje výsledky a zapojiť sa tým do rebríčku, kde víťazí tím s najlepšimi výsledkami. Každý tím má obmedzený počet požiadaviek na testovanie výsledkov, aby sa zabránilo prispôsobovaniu výsledkov na testovaciu množinu. Táto množina neobsahuje výsledné obrázky, tie sú privátne uložené na VUT v Brne a používajú sa na porovnanie s výslednými obrázkami modelov, pričom sa následne vypočíta skóre modelu pomocou niekoľkých metrík. Na to bola vytvorená celá MLFDB stránka¹, na ktorej má ktokoľvek možnosť nahrať dosiahnuté výsledky a nechať ich objektívne vyhodnotiť [38].

¹<http://splab.cz/mlfdb/results.html>

4 Záver

Práca sa zaoberá jednou z dôležitých oblastí určovania podobnosti objektov na základe obrazovej informácie – biometrickou identifikáciou osôb pomocou tváre. Práca sa primárne zameriava na výskum superrozlíšenia tváre zo sekvencie obrázkov nízkeho rozlíšenia (nekvalitné video) a porovnáva tento prístup s jednosnímkovými metódami, ktoré sú stále považované za najpresnejšie v tejto oblasti. Aj napriek výraznému technologickému pokroku a automatizovaným metódam je tu stále priestor pre zlepšovanie presnosti rozpoznávacích algoritmov, ktoré zlyhávajú predovšetkým pri hraničných situáciách, kedy je ešte možné osoby rozpoznať.

Riešenie práce začína od všeobecnejších problémov, ktoré však stavajú na rovnakých základoch (strojové učenie, obrazové spracovanie dát) a postupne sa prechádza ku hlavnému cieľu. Funkčnosť navrhnutých metód bola overená na vybraných problémoch a výsledky boli publikované na medzinárodných konferenciách: klasifikácia ručne písaného textu pomocou tradičných metód strojového učenia [4] a následne pomocou hlbokého učenia [2], detekcia hypomímie z obrázku tváre [30], vplyv šumu obrázku na trénovanie [34, 35], detekcia osôb [58]. Na národnej konferencii bol predstavený škálovateľný systém schopný bežať v reálnom čase [32], pričom kombináciou špeciálneho hardware a unikátneho software vznikol prototyp systému biometrickej identifikácie [59]. V [6] (Q1, IF: 3,998) je popísaná klasifikácia materiálu zo spektroskopie algoritmom SVM (presnosť 99,2 % pre 13 hornín). Výskum čiastočne zasahoval aj do oblasti spracovania prirodzeného jazyka, tieto výstupy boli spracované ako kapitola knihy „Big Data for Service operations Management“ vydavateľstva Springer (čaká na publikovanie).

Hlavný prínos práce je návrh metód pre superrozlíšenie tváre zo sekvencie snímok nízkeho rozlíšenia. Jedná sa predovšetkým o prípady, kde identifikácia osôb je na hranici rozpoznania. Pre overenie navrhnutých metód bola vytvorená nová trénovala a testovacia množina obsahujúca 17426 sekvencií tvárí nízkeho rozlíšenia [38]. Veľkosťou je porovnateľná s poprednými svetovými databázami, avšak oproti stávajúcim databázam neobsahuje iba známe osobnosti, ale zahŕňa rôznych ľudí, rôzne etniká, obe pohlavia a širokú škálu veku, a obsahuje snímky z odlišných snímacích zariadení, nasvietení, uhlov apod. Navrhnuté metódy prekonali výkonnosť jednosnímkových metód, ktoré sú stále považované za najlepšie v súčasnosti, konkrétne najlepšie výsledky dosahuje metóda U-Net+GEU3. Metódy boli porovnávané pomocou objektívnych metrík a ľudského hodnotenia, pričom bolo dokázané, že výsledky objektívnych metrík neodpovedajú vnímaniu človeka. Porovnanie týchto metód potvrdilo hypotézu, že viacsnímkové metódy majú lepšiu schopnosť kvalitnejšie vytvoriť výstup ako jednosnímkové metódy. Použité architektúry, dataset a zdrojové kódy boli zverejnené spolu s vybranými výstupmi v časopise (Q2, IF: 2,474) v [38].

Literatúra

- [1] Bartošová, I. K.; Třečková, E.: The New Writing System Comenia Script in the Czech Republic. *Procedia-Social and Behavioral Sciences*, ročník 112, 2014: s. 1255–1262.
- [2] Rajnoha, M.; Burget, R.; Dutta, M. K.: Handwriting comenia script recognition with convolutional neural network. In *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2017, s. 775–779.
- [3] Smith, R.: An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, ročník 2, IEEE, 2007, s. 629–633.
- [4] Rajnoha, M.; Burget, R.; Dutta, M. K.: Offline handwritten text recognition using support vector machines. In *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2017, s. 132–136.
- [5] Povoda, L.; Burget, R.; Masek, J.; aj.: Optimization Methods in Emotion Recognition System. *Radioengineering*, ročník 25, č. 3, 2016: s. 565–572.
- [6] Pořízka, P.; Klus, J.; Mašek, J.; aj.: Multivariate classification of echellograms: a new perspective in Laser-Induced Breakdown Spectroscopy analysis. *Scientific Reports*, ročník 7, č. 1, 2017: s. 1–12.
- [7] Laughlin, S.: A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung c*, ročník 36, č. 9-10, 1981: s. 910–912.
- [8] Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; aj.: LIBLINEAR: A library for large linear classification. *Journal of machine learning research*, ročník 9, č. Aug, 2008: s. 1871–1874.
- [9] Chollet, F.: Deep learning library for theano and tensorflow, <https://keras.io>. 2015.
- [10] Robbins, H.; Monro, S.: A stochastic approximation method. *The annals of mathematical statistics*, 1951: s. 400–407.
- [11] Prechelt, L.: Early stopping-but when? In *Neural Networks: Tricks of the trade*, Springer, 1998, s. 55–69.
- [12] Liu, W.; Anguelov, D.; Erhan, D.; aj.: Ssd: Single shot multibox detector. In *European conference on computer vision*, Springer, 2016, s. 21–37.

- [13] Everingham, M.; Van Gool, L.; Williams, C. K.; et al.: The PASCAL visual object classes challenge 2007 (VOC2007) results. 2007.
- [14] Torrey, L.; Shavlik, J.: Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, 2010, s. 242–264.
- [15] Kingma, D. P.; Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [16] Koch, G.; Zemel, R.; Salakhutdinov, R.: Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, ročník 2, Lille, 2015.
- [17] Rajnoha, M.: Realtime Pedestrian Recognition Using Siamese Network. In *In Proceedings of the 24th Conference STUDENT EEICT 2018*, Brno University of Technology, Faculty of Electrical Engineering and Communication, 2018, s. 441–445.
- [18] Fihl, P.: AVSS 2007: IEEE International Conference on Advanced Video and Signal based Surveillance, London, UK, September 2007: Conference participation. 2009.
- [19] Jain, V.; Learned-Miller, E.: Fddb: A benchmark for face detection in unconstrained settings. Technická zpráva, UMass Amherst technical report, 2010.
- [20] Tangsen, P.; Wu, Z.; Yamaguchi, K.: Looking at outfit to parse clothing. *arXiv preprint arXiv:1703.01386*, 2017.
- [21] The BioID Face Database, <https://www.bioid.com/>.
- [22] Hancock, P.: Psychological image collection at stirling (pics), <http://pics.psych.stir.ac.uk>. 2008.
- [23] Redmon, J.; Farhadi, A.: Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [24] Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [25] Geitgey, A.: face-recognition, GitHub, https://github.com/ageitgey/face_recognition.
- [26] Schroff, F.; Kalenichenko, D.; Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, s. 815–823.

- [27] Huang, G. B.; Mattar, M.; Berg, T.; aj.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. 2008.
- [28] Geitgey, A.: Machine learning is fun! part 4: modern face recognition with deep learning. *Medium. Medium Corporation*, ročník 24, 2016.
- [29] Kazemi, V.; Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, s. 1867–1874.
- [30] Rajnoha, M.; Mekyska, J.; Burget, R.; aj.: Towards Identification of Hypomimia in Parkinson’s Disease Based on Face Recognition Methods. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, s. 1–4.
- [31] Dalal, N.; Triggs, B.: Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, ročník 1, IEEE, 2005, s. 886–893.
- [32] Rajnoha, M.: Scalable Person Identification System for Real-time Applications. In *In Proceedings of the 25th Conference STUDENT EEICT 2019*, Brno University of Technology, Faculty of Electrical Engineering and Communication, 2019, s. 500–504.
- [33] Reijnders, J. S.; Ehrt, U.; Weber, W. E.; aj.: A systematic review of prevalence studies of depression in Parkinson’s disease. *Movement Disorders*, ročník 23, č. 2, 2008: s. 183–189.
- [34] Rajnoha, M.; Mikulec, V.; Burget, R.; aj.: A Perspective of the Noise Removal for Faster Neural Network Training. In *2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2019, s. 1–4.
- [35] Rajnoha, M.; Burget, R.; Povoda, L.: Image background noise impact on convolutional neural network training. In *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2018, s. 1–4.
- [36] Viola, P.; Jones, M. J.: Robust real-time face detection. *International journal of computer vision*, ročník 57, č. 2, 2004: s. 137–154.
- [37] Isola, P.; Zhu, J.-Y.; Zhou, T.; aj.: Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, s. 1125–1134.

- [38] Rajnoha, M.; Mezina, A.; Burget, R.: Multi-Frame Labeled Faces Database: Towards Face Super-Resolution from Realistic Video Sequences. *Applied Sciences*, ročník 10, č. 20, 2020: str. 7213.
- [39] Salvetti, F.; Mazzia, V.; Khaliq, A.; aj.: Multi-Image Super Resolution of Remotely Sensed Images Using Residual Attention Deep Neural Networks. *Remote Sensing*, ročník 12, č. 14, 2020: str. 2207.
- [40] Molini, A. B.; Valsesia, D.; Fracastoro, G.; aj.: DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images. *IEEE Transactions on Geoscience and Remote Sensing*, ročník 58, č. 5, 2019: s. 3644–3656.
- [41] Deudon, M.; Kalaitzis, A.; Goytom, I.; aj.: HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery. *arXiv preprint arXiv:2002.06460*, 2020.
- [42] Dong, C.; Loy, C. C.; He, K.; aj.: Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, ročník 38, č. 2, 2015: s. 295–307.
- [43] Lim, B.; Son, S.; Kim, H.; aj.: Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, s. 136–144.
- [44] Ledig, C.; Theis, L.; Huszár, F.; aj.: Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, s. 4681–4690.
- [45] Wang, X.; Yu, K.; Wu, S.; aj.: Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, s. 0–0.
- [46] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, s. 234–241.
- [47] He, K.; Zhang, X.; Ren, S.; aj.: Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, s. 770–778.
- [48] Bare, B.; Yan, B.; Ma, C.; aj.: Real-time video super-resolution via motion convolution kernel estimation. *Neurocomputing*, ročník 367, 2019: s. 236–245.

- [49] Shi, W.; Caballero, J.; Huszár, F.; aj.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, s. 1874–1883.
- [50] Mohammadi, P.; Ebrahimi-Moghadam, A.; Shirani, S.: Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*, 2014.
- [51] Poobathy, D.; Chezian, R. M.: Edge detection operators: Peak signal to noise ratio based comparison. *IJ Image, Graphics and Signal Processing*, ročník 10, 2014: s. 55–61.
- [52] Wang, Z.; Bovik, A. C.; Sheikh, H. R.; aj.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, ročník 13, č. 4, 2004: s. 600–612.
- [53] Narvekar, N. D.; Karam, L. J.: A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In *2009 International Workshop on Quality of Multimedia Experience*, IEEE, 2009, s. 87–91.
- [54] Ferzli, R.; Karam, L. J.: A no-reference objective image sharpness metric based on just-noticeable blur and probability summation. In *2007 IEEE International Conference on Image Processing*, ročník 3, IEEE, 2007, s. III–445.
- [55] Ferzli, R.; Karam, L. J.: A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE transactions on image processing*, ročník 18, č. 4, 2009: s. 717–728.
- [56] Krizhevsky, A.; Sutskever, I.; Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012, s. 1097–1105.
- [57] Kolář, M.; Hradiš, M.; Zemčík, P.: Deep learning on small datasets using on-line image search. In *Proceedings of the 32nd Spring Conference on Computer Graphics*, 2016, s. 87–93.
- [58] Rajnoha, M.; Povoda, L.; Masek, J.; aj.: Pedestrian detection from low resolution public cameras in the wild. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2018, s. 291–295.
- [59] Rajnoha, M.; Burget, R.; Rainish, J.: CamBrain; Prototyp systému pro monitorování hledaných osob v reálném čase. Technická zpráva, Brno University of Technology, 2020.

Curriculum Vitae

Martin Rajnoha

Osobné informácie

Dátum narodenia: 30.5. 1991
Miesto narodenia: Bojnice
Telefón: +420 770 127 942
E-mail: rajnohamato@gmail.com

Vzdelanie

2016 – 2020 Doktorské štúdium – Teleinformatika
Fakulta elektrotechniky a komunikačných technológií
Vysoké učení technické v Brně, titul: Ph.D. (predp. uk. 2020)
2013 – 2016 Magisterské štúdium – Telekomunikační a informační technika
Fakulta elektrotechniky a komunikačních technologií
Vysoké učení technické v Brně, titul: Ing.
2010 – 2013 Bakalárske štúdium – Teleinformatika
Fakulta elektrotechniky a komunikačních technologií
Vysoké učení technické v Brně, titul: Bc.
2006 – 2010 Úplné stredné odborné vzdelanie - Mechanik počítačových sietí
Stredná odborná škola Handlová

Stáže

apríl – 2019 University of Maribor, Slovinsko, program Ceepus
nov. – 2018 University of Alberta, Edmonton, Kanada, stáž
apríl – 2018 University of Maribor, Slovinsko, program Ceepus
2015 – 2016 Technical University of Sofia, Bulharsko, program Erasmus
2013 – 2014 Bogazici University, Istanbul, Turecko, program Erasmus

Ocenenia

- 2019 – 1. miesto na študentskej konferencii EEICT
- 2018 – „Best paper award“ v sekcii konferencie ICUMT, Moskva
- 2016 – Cena dekanky za diplomovú prácu
- 2016 – 2. miesto na študentskej konferencii EEICT
- 2010 – Cisco NetAcad games, 2. miesto, medzinárodné kolo, kategória tímy
- 2009 – Cisco NetAcad games, 2. miesto, medzinárodné kolo, kategória tímy
- 2008 – CCNA certifikát

Zamestnanie

2018 – súč.	Vývojár, datový analytik Profitak solutions s.r.o.; Mínská 13, Brno
2016 – 2019	Vedecký pracovník Centra senzorických informačních a komunikačních systémů Vysoké učení technické v Brně; Technická 3058/10, Brno
2015 – 2016	Software tester EmbedIT; Holandská 10, Brno

Spolupráca na projektoch

2019 – 2020	Společný výzkum zdravotní diagnostiky a pokročilé analýzy oftalmologických medicínských dat University of Science and Technology Beijing, reg. č. 8JCH1067
2017 – 2020	Expertní systém pro automatickou analýzu a řízení big data úložišť výrobních společností MPO reg. č. FV20044
2017 – 2019:	Hardwarově akcelerovaná identifikace osob z videozáznamů s použitím hlubokého strojového učení MVČR reg. č. VI2VS/554
2017 – 2019:	Multimodální zpracování nestrukturovaných dat s využitím strojového učení a sofistikovaných metod analýzy signálů a obrazů reg. č. FEKT-S-17-4476
2016 – 2016:	Connected homes - thermostat datamining Hospodářska zmluva (750 669 CZK), referenčná osoba - Ondřej Lipták - Leader of electronic design group at Honeywell

Vyžiadané prednášky

- 2019 – University of Science and Technology Beijing, Čína
- 2018 – University of Alberta, Edmonton, Kanada

Publikačná aktivita

- Publikácie v časopisoch s impakt faktorom: **2**
- Publikácie kapítol v knihách: **1** (priaté 23.11. 2020, čaká na publikovanie)
- Publikácie v časopisoch bez impakt faktoru: **1**
- Publikácie v konferenčných zborníkoch: **14**
- Prototyp: **1**, Software: **2**
- Príspevky indexované v databázi WoS: **10**, Scopus: **10**
- H-index podľa databázi WoS: **3**, Scopus: **3**

ABSTRAKT

Vzhľadom na meniacu sa bezpečnostnú situáciu vo svete, monitorovanie verejných priestorov a jeho automatické spracovanie v reálnom čase sa stáva čoraz významnejšie. Problémom je však analýza nekvalitných záznamov, pričom v týchto prípadoch aj najmodernejšie vedecké metódy zlyhávajú. Táto práca skúma jednu z dôležitých oblastí podobnosti obrazov – biometrickú identifikáciu osôb pomocou tváre. Práca sa primárne zaoberá superrozlíšením tváre zo sekvencie obrázkov nízkeho rozlíšenia a porovnáva tento prístup s jednosnímkovými metódami, ktoré sú stále považované za najpresnejšie. Pre tento účel bol vytvorený nový trénovací dataset, určený priamo pre viacsnímkové superrozlíšenie tvárí z nízkej kvality vstupnej sekvencie, ktorý je veľkosťou zrovnateľný s poprednými svetovými datasetmi. Výsledky boli zhodnotené pomocou prieskumu ľudského vnímania, ale aj podľa presne definovaných objektívnych metrík. Porovnanie týchto metód potvrdilo hypotézu, že viacsnímkové metódy dosahujú lepšie výsledky ako jednosnímkové metódy. Použité architektúry a zdrojové kódy boli zverejnené spolu s vytvoreným datasetom, čím sa vytvoril základ pre budúci výskum v tejto oblasti.

ABSTRACT

Monitoring of public areas and their automatic real-time processing became increasingly significant due to the changing security situation in the world. However, the problem is an analysis of low-quality records, where even the state-of-the-art methods fail in some cases. This work investigates an important area of image similarity – biometric identification based on face image. The work deals primarily with the face super-resolution from a sequence of low-resolution images and it compares this approach to the single-frame methods, that are still considered as the most accurate. A new dataset was created for this purpose, which is directly designed for the multi-frame face super-resolution methods from the low-resolution input sequence, and it is of comparable size with the leading world datasets. The results were evaluated by both a survey of human perception and defined objective metrics. A hypothesis that multi-frame methods achieve better results than single-frame methods was proved by a comparison of both methods. Architectures, source code and the dataset were released. That caused a creation of the basis for future research in this field.