

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2018

Veronika Horňáková



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

ROZPOZNÁNÍ TVÁŘÍ V REÁLNÝCH PODMÍNKÁCH POMOCÍ TECHNIK HLUBOKÉHO UČENÍ

DEEP LEARNING BASED FACE RECOGNITION IN REAL CONDITIONS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Veronika Horňáková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Přinosil, Ph.D.

BRNO 2018

Bakalářská práce

bakalářský studijní obor **Teleinformatika**
Ústav telekomunikací

Studentka: Veronika Horňáková

ID: 189507

Ročník: 3

Akademický rok: 2017/18

NÁZEV TÉMATU:

Rozpoznání tváří v reálných podmínkách pomocí technik hlubokého učení

POKYNY PRO VYPRACOVÁNÍ:

Prostudujte stávající stav v oblasti rozpoznání lidských tváří v obrazovém signálu pomocí technik hlubokého učení. Vytipujte několik vhodných existujících řešení a proveďte jejich instalaci a zprovoznění. Na základě testování nad reálnými daty stanovte přesnost a časovou náročnost jednotlivých přístupů. Dále navrhnete vlastní metodu pro rozpoznání osob ve videosekvencích v případě, že máte k dispozici více než jeden referenční snímek dané osoby. Navržený přístup ověřte na datech z reálného prostředí.

DOPORUČENÁ LITERATURA:

- [1] BALTRUŠAITIS, Tadas; ROBINSON, Peter; MORENCY, Louis-Philippe. Openface: an open source facial behavior analysis toolkit. In: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016. 1-10.
- [2] SCHROFF, Florian; KALENICHENKO, Dmitry; PHILBIN, James. Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. 815-823.

Termín zadání: 5.2.2018

Termín odevzdání: 29.5.2018

Vedoucí práce: Ing. Jiří Přinosil, Ph.D.

Konzultant:

prof. Ing. Jiří Mišurec, CSc.
předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Táto bakalárska práca sa zaoberá rozpoznávaním ľudských tvárí, pomocou techník hlbokého učenia. Rozpoznávanie tvárí sa používa pre dve hlavné úlohy a to overenie a identifikácia. V práci sú opísané techniky hlbokého učenia najmä konvolučné neurónové siete, ktoré sú najvýznamnejšou metódou pre spracovanie obrazov – detekcia, klasifikácia a segmentácia obrazu. Proces rozpoznávania tváre prebieha v štyroch hlavných krokoch ktoré sú: detekcia tváre, výber tváre, extrakcia tváre a klasifikácia tváre. Boli vytipované tri existujúce riešenia rozpoznávania tváre (OpenFace, FaceNet a Face_Recognition), ktoré sú popísané v práci, najmä princíp na ktorom rozpoznávajú ľudské tváre. Pomocou testovania s dátovou sadou Labeled Faces in the Wild (LFW), bola stanovená presnosť a časová náročnosť jednotlivých aplikácií. Testovanie FaceNet a Face_Recognition prebehlo aj na reálnych dátach s detekciou tvárí vo videu a pri sťažených podmienkach. Test sa zaoberá porovnaním dvojice obrázkov a určením či sa jedná o tú istú osobu. Výsledky testovania sú zobrazené v prehľadnom grafe a tabuľke.

KLÚČOVÉ SLOVÁ

FaceNet, Face_Recognition, hlboké učenie, konvolučná neurónová sieť, LFW, OpenFace, rozpoznávanie tvárí, XGBoost.

ABSTRACT

This bachelor thesis explores the area of face recognition using deep learning technique. Face recognition is used for two main reasons: verification and identification. In this thesis we describe the techniques of deep learning, mostly the convolutional neural networks, which are the most significant method for processing images – detection, classification and segmentation of the image. The process of face recognition is divided into four main steps: face detection, face selection, face extraction and face classification. We chosen three of the existing programs for face recognition (OpenFace, FaceNet and Face_Recognition), which are described in this thesis, in particular the principle of the human face recognition. Thanks to the tests with the data set of Labeled Faces in the Wild (LFW) we could specify the accuracy and the time requirement of each application. Testing of FaceNet and Face_Recognition ran on real data with face detection in video with complicated conditions. The test compares two images and tries to determine if is the same person. The test results are show in graph and table.

KEYWORDS

FaceNet, Face_Recognition, deep learning, convolutional neural networks, LFW, OpenFace, facial recognition, XGBoost.

HORŇÁKOVÁ, Veronika. *Rozpoznání tváří v reálných podmínkách pomocí technik hlbokého učení*. Brno, 2018, 42 s. Bakalárska práca. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedúci práce: Ing. Jiří Přinosil, Ph.D.

VYHLÁSENIE

Vyhlasujem, že som svoju bakalársku prácu na tému „Rozpoznání tváří v reálných podmínkách pomocí technik hlubokého učení“ vypracovala samostatne pod vedením vedúceho bakalárskej práce, využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autorka uvedenej bakalárskej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto bakalárskej práce som neporušila autorské práva tretích osôb, najmä som nezasiahla nedovoleným spôsobom do cudzích autorských práv osobnostných a/alebo majetkových a som si plne vedomá následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona Českej republiky č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon), v znení neskorších predpisov, vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákoníka Českej republiky č. 40/2009 Sb.

Brno

.....

podpis autora(-ky)

POĎAKOVANIE

Rada by som poďakovala môjmu vedúcemu bakalárskej práce pánovi Ing. Jiřímu Přinosilovi, Ph.D. za odborné vedenie, konzultácie, trpezlivosť a podnetné návrhy k práci.

Brno

.....

podpis autora(-ky)

OBSAH

Úvod	9
1 Rozpoznávanie ľudských tvárí	10
1.1 Architektúra hlbokého učenia	11
1.1.1 Konvolučné neurónové siete	12
1.1.2 Hlboké neurónové siete	14
2 Aplikácie rozpoznávania tváre	15
2.1 OpenFace	15
2.2 FaceNet	18
2.3 Face_Recognition	20
3 Experimentálna časť	23
3.1 Databáza Labeled Faces in the Wild:	23
3.2 Testovanie s programom XGBoost	24
3.2.1 Cieľ a funkcia	24
3.2.2 Výhody XGBoostu	25
4 Výsledky experimentálnej časti	27
4.1 Výsledky testovania s databázou LFW	27
4.2 Výsledky testu s databázou videí a obrázkov	30
4.3 Výsledky testovania s programom XGBoost	33
Záver	36
Literatúra	38
Zoznam príloh	41
A Obsah priloženého CD	42

ZOZNAM OBRÁZKOV

1.1	Architektúra systému rozpoznávania tváre [2].	10
1.2	Digitálna reprezentácia obrazu a matica konvolúcie [9].	12
1.3	Príklad výpočtu konvolúcie [9].	13
1.4	Model hlbkej neurónovej siete pre rozpoznávanie výrazu tváre [4].	14
2.1	OpenFace: analýza tváre vrátane detekcie orientačných bodov na tvári, sledovanie pohybu hlavy, sledovanie očného pohľadu a rozpoznanie akčnej jednotky [10].	15
2.2	Vzor odhadu pohľadu, zelené čiary predstavujú odhadované vektory pre očný pohľad [10].	16
2.3	Logický tok pre rozpoznávanie tváre s neurónovými sieťami [11].	17
2.4	Štruktúra OpenFace projektu [11].	17
2.5	Osvetlenie a variácie postojov [13].	18
2.6	Tréningová sieť na rozlíšenie troch obrázkov [15].	19
2.7	Obrys očí, nosa, úst a brady [16].	20
2.8	Nahradenie pixelov gradientami [18].	21
2.9	Transformácia tváre tak, aby bola dokonale centrovaná [18].	22
3.1	Výsledky testovania rýchlosti[21].	25
4.1	Závislosť medzi FP a citlivosťou.	28
4.2	Nesprávne určenie rozdielnych osôb ako rovnakých – Face_Recognition.	28
4.3	Nesprávne určenie rozdielnych osôb ako rovnakých – FaceNet.	29
4.4	Vzorové obrázky z databáze s videami.	30
4.5	Vykreslenie normovaných váh pre 3 osoby	32
4.6	Závislosť medzi FP pomerom a citlivosťou	34
4.7	Počet videí určených bez chyby, s chybou a neurčených vôbec	35
4.8	Percentuálne vyjadrenie správnych a nesprávnych snímok vo všetkých videách	35

ZOZNAM TABULIEK

4.1	Výsledky testovania.	27
4.2	Výsledky testovania na databáze s videami	30
4.3	Výsledky testovania s použitím priemeru a mediánu.	31
4.4	Výsledky testovania s použitím priemeru a normovaných váh	32
4.5	Výsledky testovania XGBoostu	33

ÚVOD

Táto práca sa zaoberá rozpoznávaním ľudských tvárí pomocou techník hlbokého učenia. Stále častejšie sa stretávame s implementáciami, ktoré dovoľujú rozpoznávať ľudí medzi sebou.

Doposiaľ sa nepodarilo vyvinúť systém s presnosťou rozpoznávania tvárí 100%. Existujú už však systémy na rozpoznávanie tvárí, ktoré sa blížia k tejto hranici. Napr. FaceNet od spoločnosti Google dosahuje úspešnosť cez 99%.

Táto práca porovnáva stav súčasných systémov na rozpoznávanie ľudských tvárí z hľadiska presnosti a časovej náročnosti.

Prvá časť sa zaoberá teóriou rozpoznávania ľudských tvárí a implementáciou hlbokého učenia. Sú tu popísané hlavné úlohy rozpoznávania tvárí. Túto časť tvorí popis architektúry hlbokého učenia a rozdelenie na konvolučné neurónové siete a hlboké neurónové siete.

V druhej časti sú podrobne popísané vytipované existujúce riešenia na rozpoznávanie ľudských tvárí. Z existujúcich riešení bol zvolený FaceRecognition, ktorý je považovaný za najjednoduchšiu knižnicu rozpoznávania tvárí, FaceNet ktorý sa vyznačuje dobrou presnosťou a OpenFace, ktorý bol z praktického testovania vynechaný, vzhľadom na horšiu presnosť rozpoznávania osôb, ktorú udáva na svojich internetových stránkach.

Poslednú časť tvoria výsledky testovania. Testovanie bolo zamerané na presnosť a časovú náročnosť vybraných programov. Prvé testovanie prebehlo na dátovej sade Labeled Faces in the Wild (LFW). Druhý test prebehol s reálnymi dátami a rozpoznávaním ľudí vo videu. K testovaniu sa používala databáza 24 neznámych osôb. V práci bola navrhnutá vlastná metóda pre rozpoznávanie osôb vo videosekvenciách v prípade, že existuje viac ako jeden referenčný snímok danej osoby. Navrhnutá metóda bola otestovaná a výsledky boli zverejnené v tabuľkách a grafe. Na automatizáciu rozpoznávania osôb bol použitý program XGBoost, pri ktorom si užívateľ môže vytvoriť vlastný model na testovanie. Na záver bolo otestované rozpoznávanie osôb vo videosekvenciách na jednotlivých snímkoch videí.

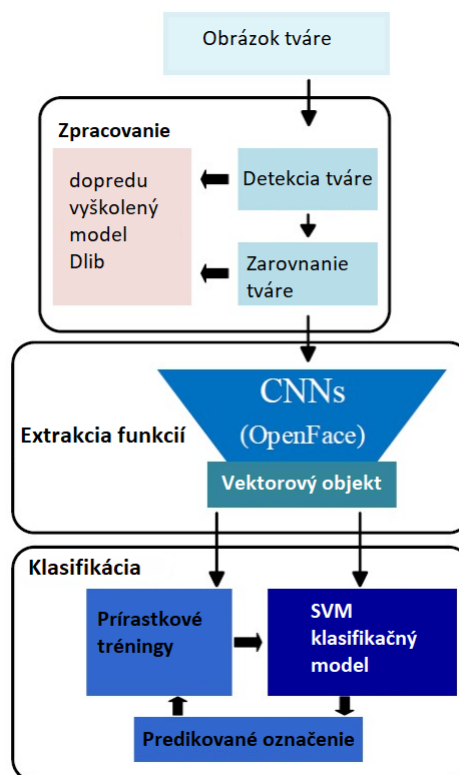
1 ROZPOZNÁVANIE ĽUDSKÝCH TVÁRÍ

Rozpoznanie tváre je aktívna téma výskumu od sedemdesiatych rokov [11].

Rozpoznávanie tvári sa používa pre dve hlavné úlohy: [1]

1. overenie (prispôsobenie jeden k jednému): pri prezentácii s obrazom tváre neznámeho človeka s tvrdením o totožnosti, zisťujúc, či jedinou osobou je ten, o ktorom tvrdí, že je;
2. identifikácia (prispôsobenie jeden na mnohé): vzhľadom k obrazu neznámeho človeka, určiť jeho osobnú identitu porovnaním s obrazcami z databáz obrázkov známych osôb.

Na obrázku 1.1 je znázornený systém, ktorý je zameraný na proces rozpoznávania tváre v štyroch hlavných krokoch: [2] detekcia tváre, výber tváre, extrakcia tváre a klasifikácia tváre.



Obr. 1.1: Architektúra systému rozpoznávania tváre [2].

Obrázky tváří musia byť odolné voči interpersonálnemu obrazu variácií [11], ako je vek, výrazy a styling. Jafra a Arabnia [1] poskytujú komplexný prehľad o technikách rozpoznávania tváre do roku 2009.

1.1 Architektúra hlbokého učenia

Hlboké učenie je podskupina strojového učenia sa, ktorá sa zameriava na algoritmy inšpirované štruktúrou a funkciou mozgu nazývanými umelé neurónové siete [3]. Okrem toho že konvolučné neurónové siete sú škálovateľné, ich ďalším prínosom je schopnosť vykonávať automatickú extrakciu prvkov z nespracovaných údajov. Yoshua Bengio je ďalším lídrom v hlbokom vzdelávaní, hoci začal so silným záujmom o automatické učenie, ktoré veľké neurónové siete dokážu dosiahnuť [3]. Hlboké učenie vyniká v problémových oblastiach, kde sú vstupy (a dokonca aj výstupy) analógové. Znamená to, že v tabuľkovom formáte nie sú len malé množstvá, ale obrázky pixelových údajov, dokumentov z textových dát alebo súborov zvukových údajov [3]. V dôsledku skúmania hlbokého učenia za posledné roky [4], existuje veľké množstvo súvisiacich prístupov. Pre spracovávanie obrazov sú najvýznamnejšie konvolučné neurónové siete. Neurónové siete nám pomáhajú zoskupovať a klasifikovať [5]. Čoraz viac spoločností využíva hlboké učenie v jadre svojich služieb [6]. Facebook využíva neurónové siete na automatické značkovanie algoritmov, Google pre vyhľadávanie fotografií, Amazon pre odporúčania produktov a Instagram pre svoju vyhľadávaciu infraštruktúru. Avšak klasický a pravdepodobne najpopulárnejší prípad používania týchto sietí je pre spracovanie obrazu. Za posledné roky nastalo veľké zlepšenie vyhľadávania obrázkov napríklad v Google vyhľadávači a takisto čoraz lepšiu identifikáciu tváří na Facebooku, ktorý ponúka „otagovanie“ osôb [7]. Konvolučné neurónové siete majú prepojenie vrstvy zatiaľ čo hlboké neurónové siete tvorí len vrstva plného spojenia. V dôsledku toho sa konvolučná neurónová sieť používa pre rozpoznávanie obrazov a hlboké neurónové siete sa obvykle používajú len pre spracovanie signálu 1D ako rozpoznávanie reči. Vrstvy sú vytvorené z neurónov, často nazývaných ako uzol alebo jednotka [8]. Uzol je miesto, kde sa uskutočňuje výpočet. Uzol kombinuje vstup z dát so sústavou koeficientov, ktoré zosilňujú alebo tlmia tento vstup, čím priradia významnosť vstupom pre úlohu, ktorú sa algoritmus snaží získať. Tieto vstupné váhy sú sumarizované a súčet prechádza cez tzv. aktivačnú funkciu uzla, aby sa určilo, či a do akej miery tento signál postupuje ďalej cez sieť a ovplyvňuje konečný výsledok, napríklad klasifikáciu. Úroveň vrstvy je riadok týchto neurónových uzlov. Výstup každej vrstvy je súčasne vstupom nasledujúcej vrstvy, počnúc vstupnou vrstvou, ktorá prijíma dáta.

1.1.1 Konvolučné neurónové siete

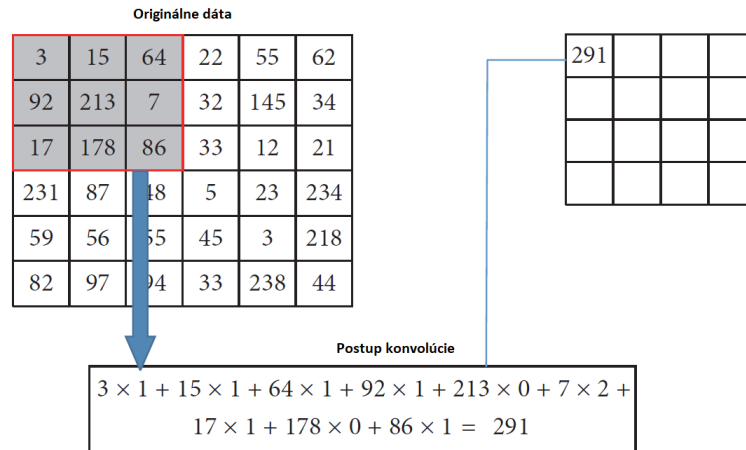
Ide o softvérové siete, ktoré pozostávajú z vrstiev malých výpočtových jednotiek, tzv. neurónov, ktoré hierarchicky posielajú spracovanú informáciu [7]. Reprezentáciu obrázka, tvoria neurónové siete tréované v rámci rozpoznávacieho procesu. Ten sa dá mierne prirovnať k postupnému zostrovaniu. Prvá vrstva určuje, kde sú na fotke svetlé a tmavé miesta, ďalšia vrstva začína rozpoznávať hrany, ďalšia tvary, nasledujúca objekty, ďalšia typy objektov a napokon dôjde k celkovému hodnoteniu situácie a pochopenia významu. Konvolučné neurónové siete sú koncepčne (ale nie exaktne) podobné biologickému systému neurónov a synapsií mozgu. Nie sú jediným riešením, sú však najrobustnejšie a používajú sa predovšetkým na úlohy, ktoré súvisia so spracovaním obrazovej informácie [7]. Konvolučné neurónové siete sú navrhnuté tak, aby spracovávali údaje, ktoré majú podobu viacnásobného pola. Sú jedným z najvýznamnejších metód pre spracovanie obrazov [9]. Konvolučné neurónové siete sú vysoko efektívne a bežne používané v aplikáciách počítačového videnia. Skladajú sa z troch typov vrstiev: konvulčná vrstva, podsvorková vrstva a vrstva združovania [9].

- Konvulčnú vrstvu tvorí digitálny obraz a matica konvolúcie, ktorá sa obvykle nazýva filter. V konvulčnej vrstve prebieha konvolúcia vstupného obrazu s konvulčnou maticou a potom sa generuje výstupný obraz. Výstupný obraz je odpoveď filtra alebo mapa filtra. Postup konvolúcie je znázornený na obr. 1.2 a 1.3

3	15	64	22	55	62
92	213	7	32	145	34
17	178	86	33	12	21
231	87	48	5	23	234
59	56	55	45	3	218
82	97	94	33	238	44

1	1	1
1	0	2
1	0	1

Obr. 1.2: Digitálna reprezentácia obrazu a matica konvolúcie [9].



Obr. 1.3: Príklad výpočtu konvolúcie [9].

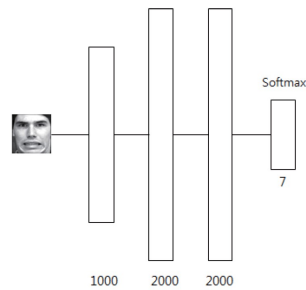
- Podsvorková vrstva znižuje veľkosť vstupného obrazu a vykonáva maximálne združovanie obrazu. Dôvodom použitia podsvorkovej vrstvy je menej parametrov čo umožňuje rýchlejšie testovanie.
- Vrstva združovania predstavuje neurónovú sieť privádzanú dopredu do vektorov s vopred definovanou dĺžkou. Neuróny v tejto vrstve majú plné spojenie so všetkými aktiváciami v predchádzajúcich vrstvách.

Týmto spôsobom konvolučné neurónové siete transformujú pôvodný obraz vrstvu po vrstve z pôvodných hodnôt pixlov na hodnotu konečnej vrstvy.

V najjednoduchšom prípade je konvolučná neurónová vrstva zoznamom vrstiev, ktoré premieňajú obrazový objem pixlov na výstupný objem. Každá vrstva akceptuje výstupný objem 3D a transformuje ju na výstupný 3D zväzok pomocou diferencovateľnej funkcie.

1.1.2 Hlboké neurónové siete

Obrázok 1.4 ukazuje hlboký neurónový model používaný v systéme rozpoznávania výrazov tváre. Vstupná snímka je zobrazená v odtieňoch šedej, pretože informácie o farbe nie sú potrebné na rozpoznanie tváre. [4] Hlboký neurónový model má tri skryté vrstvy uvedené na obrázku 1.4. Počet uzlov každej skrytej vrstvy je 1000, 2000 a 2000. Horná vrstva je softmaxová vrstva, ktorá má 7 uzlov. Každý uzol zodpovedá každej emócií. Napokon, výstup je pravdepodobnosť emócií.

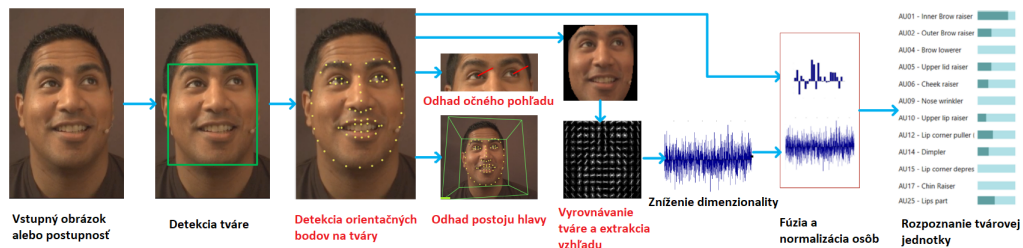


Obr. 1.4: Model hlbkej neurónovej siete pre rozpoznávanie výrazu tváre [4].

2 APLIKÁCIE ROZPOZNÁVANIA TVÁRE

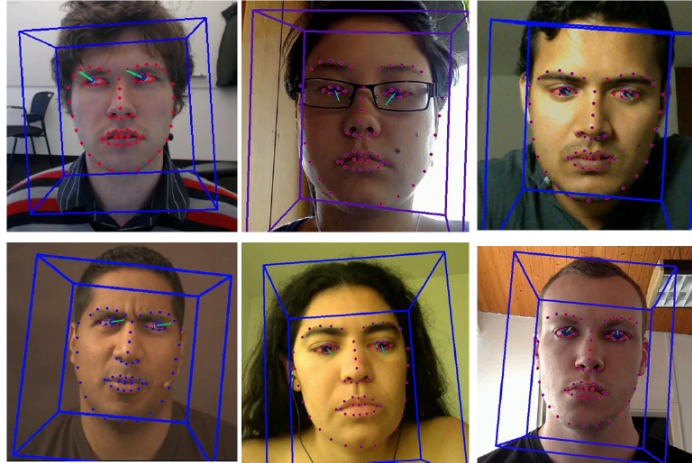
2.1 OpenFace

OpenFace je open source framework, ktorý implementuje najmodernejšie algoritmy na rozpoznávanie ľudskej tváre [10].



Obr. 2.1: OpenFace: analýza tváre vrátane detekcie orientačných bodov na tvári, sledovanie pohybu hlavy, sledovanie očného pohľadu a rozpoznanie akčnej jednotky [10].

- Detekcia orientačných bodov na tvári – existuje široký výber voľne dostupných nástrojov pre zisťovanie orientačných bodov na tvári v obrazoch alebo videách, avšak veľmi málo prístupov poskytuje zdrojový kód a namiesto toho poskytujú len spustiteľné binárne súbory. To robí reprodukciu experimentov na rôznych tréningových sadách alebo použitie rôznych orientačných bodov anotácie ťažké. Ďalej binárne súbory umožňujú iba určitú preddefinovanú funkcionálnosť a nie sú často multiplatformové, takže integrácia systému v reálnom čase, ktorá by spoľahlivo detegovala orientačné body je takmer nemožná.
- Sledovanie pohybu hlavy – v skoršom riešení odhadu pohybu hlavy je systém Watson, ktorý je implementáciou generalizovaného modelu vzhľadu založeného na AdaptiveView kamerách. Pre presné sledovanie pohybu hlavy potrebuje byť OpenFace opatrený parametrami kalibrácie kamery (ohnisková vzdialenosť, dĺžka a hlavný bod). V ich neprítomnosti OpenFace používa hrubý odhad založený na veľkosti obrázku.
- Sledovanie očného pohľadu – na sledovanie očného pohľadu využíva OpenFace rámec CLNF (Conditional Local Neural Fields). Rámec CLNF sa používa na detekciu známych orientačných bodov. Patria sem očné viečka, dúhovka a zrenica. Akonáhle je umiestnenie oka a zrenice detegované pomocou modelu CLNF, používa tieto informácie k výpočtu.



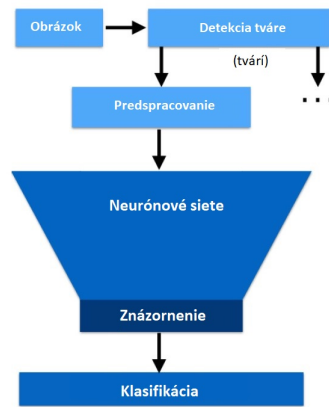
Obr. 2.2: Vzor odhadu pohľadu, zelené čiary predstavujú odhadované vektory pre očný pohľad [10].

- Detekcia akčnej jednotky – modul detekcie intenzity a prítomnosti akčnej jednotky v OpenFace je založený na najnovšom uznávanom rámci akčnej jednotky. Ide o priamu implementáciu s niekoľkými zmenami, ktoré slúžia na prispôbenie, aby lepšie fungovala na prirodzených sekvenciách videí z neviditeľných dátových súborov.

Je tiež schopný vykonávať všetky tieto úlohy spoločne v reálnom čase. Hlavné prínosy OpenFace sú [10]:

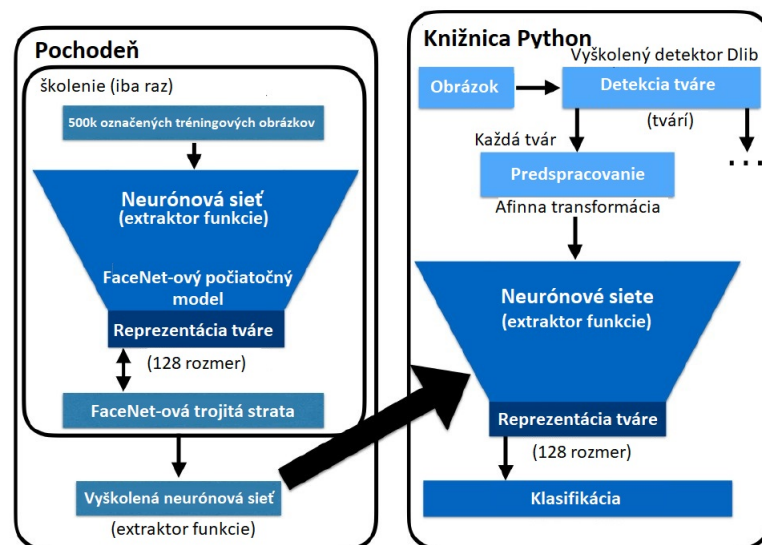
1. implementuje a rozširuje najmodernejšie algoritmy;
2. nástroj s otvoreným zdrojovým kódom;
3. prichádza s modelmi testovania, pripravenými na použitie;
4. je schopný výkonu v reálnom čase bez potreby GPU (grafického procesora, ktorý zaisťuje rýchle grafické výpočty);
5. obsahuje systém zasielania správ umožňujúcich jednoduché implementovanie interaktívnych aplikácií v reálnom čase;
6. k dispozícii ako grafické používateľské rozhranie (pre Windows) a ako nástroj príkazového riadku (pre Ubuntu, Mac OS X a Windows).

OpenFace poskytuje logický tok [11] znázornený na obrázku 2.3, aby získal nízko-rozmerné zobrazenie tváre pre tváre v obraze.



Obr. 2.3: Logický tok pre rozpoznávanie tváre s neurónovými sieťami [11].

Obrázok 2.4 znázorňuje implementáciu OpenFace. [11] Python používa knižnicu numpy pre pole a operácie lineárnej algebry, OpenCV pre primitívne počítačové videenie a scikit-learn pre klasifikáciu. Poskytuje tiež skripty, ktoré používajú matplotlib, ktorý produkuje údaje o kvalite publikácie v rôznych formátoch a interaktívnom prostredí naprieč platformami. Používa dlib – vopred vyškolený detektor tváre pre vyššiu presnosť než detektor OpenCV.



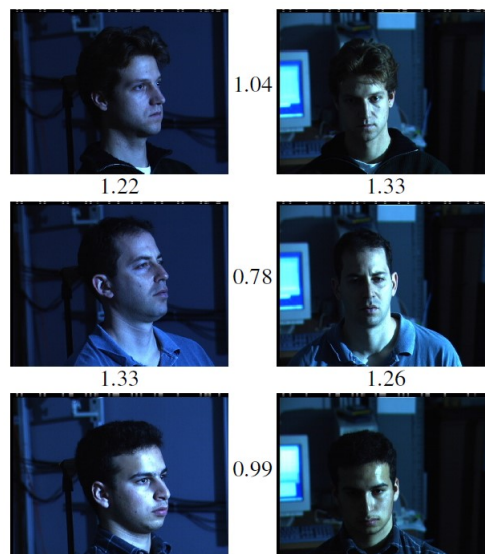
Obr. 2.4: Štruktúra OpenFace projektu [11].

Vzhľadom na to, že OpenFace udáva na svojich stránkach horšie výsledky s databázou LFW (presnosť len 0.9292 ± 0.0134) [12] ako FaceNet a FaceRecognition, som ho do svojho testovania presnosti a časovej náročnosti nezahrnula.

2.2 FaceNet

FaceNet reprezentuje jednotný systém overenia tváre (je to tá istá osoba), rozpoznania (kto je tento človek) a zoskupovania (nájdí konkrétnych ľudí medzi týmito tvármi) [13]. Sieť je nakonfigurovaná tak, aby vzdialenosti priamo zodpovedali podobnosti tváre: tváre tej istej osoby majú malé vzdialenosti a tváre odlišných ľudí majú veľké vzdialenosti. Po vytvorení tejto väzby sa stávajú vyššie uvedené úlohy jednoznačnými: overenie tváre jednoducho znamená prahovanie vzdialenosti medzi dvoma tvármi; rozpoznávanie sa stáva klasifikáciou k-NN problému a zoskupovanie sa môže dosiahnuť použitím mimoriadnych techník, ako je k-prostriedok alebo aglomeračne zhlukovanie [13]. FaceNet udáva presnosť rozpoznávania tvári s databázou LFW $0,992 \pm 0,003$ [14].

Spôsob, ktorý využíva FaceNet na rozpoznávanie tvári je čisto dátovo riadený spôsob, ktorý robí svoje zobrazenie priamo z pixelov tváre. Namiesto toho, aby používal navrhnuté funkcie, používa rozsiahlu sadu údajov označení tvári na dosiahnutie vhodnej pozície, osvetlenia a iných variačných podmienok. Obrázok 2.5 ukazuje neveriteľnú variabilitu, ktorú môže FaceNet zvládnuť.



Obr. 2.5: Osvetlenie a variácie postojov [13].

Tento obrázok zobrazuje výstupné vzdialenosti FaceNetu medzi dvojicami vzdialeností rovnakej a inej osoby v rôznych pózach a osvetľovacích kombináciách. Vzdialenosť 0:0 znamená, že tváre sú identické, 4:0 zodpovedá tomu, že tváre sú rozdielne. Môžeme vidieť, že hranica 1,1 by správne rozpoznala zhodný pár.

FaceNet priamo smeruje výstup tak, aby bol kompaktným 128-D vkladáním pomocou funkcie trojnásobnej straty založenej na klasifikácii najbližšieho blízkeho suseda (LMNN) [13]. Trojica pozostáva z dvoch zodpovedajúcich miniatúr tváre a miniatúry iného človeka. Straty majú za cieľ oddeliť pozitívnu dvojicu od zápornej vzdialenosti.



Obr. 2.6: Tréningová sieť na rozlíšenie troch obrázkov [15].

FaceNet používa hlbokú konvolučnú sieť. Využíva dve rôzne sieťové architektúry, ktoré mali veľký úspech [13].

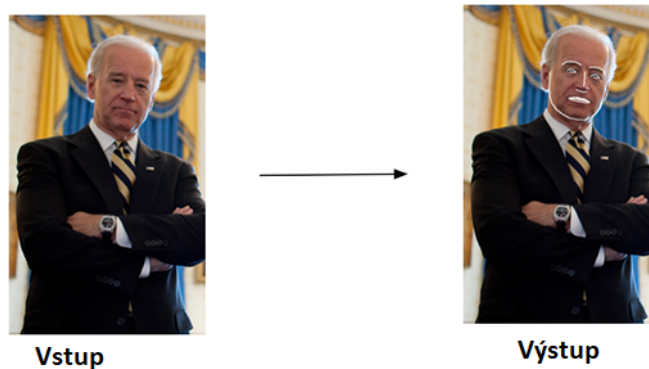
Prvá architektúra je založená na Modely Zeiler Fergus, ktorý pozostáva z niekoľkých prekladaných vrstiev záhybov, nelineárnej aktivácie, lokálnej minimalizácie odozvy a maximalizovania vrstvy združovania.

Druhá architektúra je založená na úvodnom modele Szegedy et al., ktorý bol nedávno použitý ako víťazný prístup pre ImageNet. Tieto modely môžu znížiť počet parametrov až 20-krát a majú potenciál aby znížili počet FLOPS požadovaných pre porovnateľný výkon. Výhodou FaceNetu je oveľa vyššia efektívnosť reprezentácie: dosahuje najmodernejšiu výkonnosť pri rozpoznávaní tváre pomocou iba 128 bajtov na tvár [13].

2.3 Face_Recognition

Face_Recognition sa považuje za najjednoduchšiu knižnicu rozpoznávania tvárí na svete [16]. Tento nástroj mapuje obraz ľudskej tváre do 128 dimenzionálneho vektorového priestoru, kde sú obrázky tej istej osoby blízko pri sebe a obrázky rôznych ľudí sú ďaleko od seba. Pri použití prahu vzdialenosti 0,6 úvádza model dlib presnosť 99,38 % na štandardnom referenčnom rozpoznávaní tváre LFW [17]. Táto presnosť znamená, že keď sa zobrazí dvojica obrázkov tváre, program správne identifikuje, či dvojica obrázkov patrí rovnakej alebo dvom rozdielnym osobám. Vlastnosti Face_Recognition sú:

- nájde všetky tváre, ktoré sa zobrazujú na obrázku;
- získa miesta a obrisy očí, nosa, úst a brady každého človeka;



Obr. 2.7: Obrisy očí, nosa, úst a brady [16].

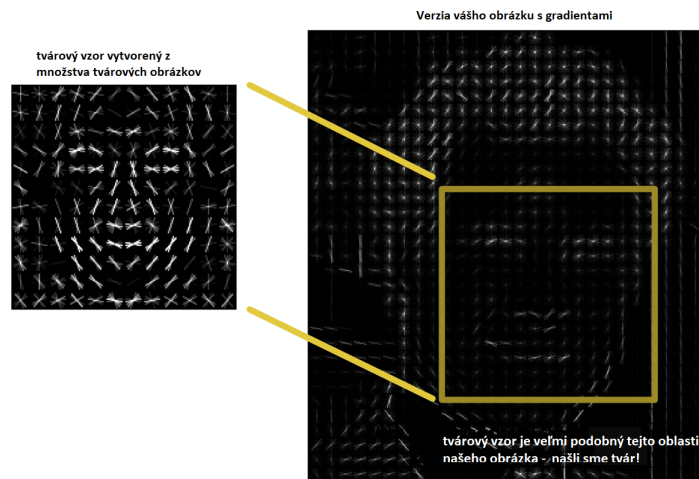
- rozpozná, kto sa nachádza na obrázku.

Prvým krokom pre rozpoznávanie tvárí pomocou Face_Recognition je **vyhľadanie všetkých tvárí** [18]. Je zrejmé, že musíme nájsť tváre na fotografii predtým, ako s nimi chceme začať pracovať. Detekcia tváre je skvelá funkcia pre fotoaparáty. Keď fotoaparát dokáže automaticky vyberať tváre, môže pred odľahčením zaistiť, že všetky tváre sú zaostrené.

Pre nájdenie všetkých tvárí v obraze stačí, že obrázok bude čierne-biele, pretože nepotrebujeme farebné údaje na vyhľadanie tvárí. Potom sa pozrieme na každý pixel v našom obraze jeden po druhom. Pri každom pixeli sa chceme pozrieť na pixely, ktoré ho priamo obklopujú. Naším cieľom je zistiť, ako tmavý je aktuálny pixel v porovnaní s okolitými pixelmi. Potom nakreslíme šípku ukazujúcu, v akom smere sa obraz stáva tmavší. Ak tento proces zopakujeme pre každý pixel v obraze, skončíme tým, že každý pixel bude nahradený šípkou. Tieto šípky sa nazývajú gradienty a ukazujú tok svetla na tmú v celom obraze. Ak analyzujeme priamo pixely, naozaj tmavé obrázky a naozaj svetlé obrázky tej istej osoby budú mať úplne odlišné

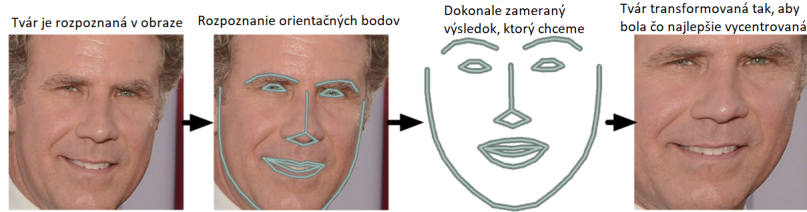
pixelové hodnoty. Ale keď zoberieme do úvahy iba smer, ktorým sa mení jas, oba skutočne tmavé obrazy a naozaj jasné obrazy skončia s rovnakým presným vyobrazením.

Ukladanie gradientu pre každý pixel dáva však príliš veľa detailov. Bolo by lepšie, keby sme mohli vidieť základný tok svetla/tmy na vyššej úrovni, aby sme mohli vidieť základný vzor obrazu. Za týmto účelom rozdelíme obrázok na malé štvorčeky s rozmermi 16×16 pixelov. Na každom štvorčeku spočítame, koľko gradientov ukazuje v každom hlavnom smere (koľko bodov smerom nahor, doprava, dole, ...). Potom nahradíme štvorec šípkou, ktorá bola zastúpená najčastejšie. Konečným výsledkom je, že pôvodný obrázok otočíme do veľmi jednoduchého zobrazenia, ktoré jednoduchým spôsobom zachytí základnú štruktúru tváre. Pomocou tejto techniky teraz môžeme ľahko nájsť tváre v akomkoľvek zobrazení.



Obr. 2.8: Nahradenie pixelov gradientami [18].

Ďalším krokom Face_Recognition je **projektovanie tváří**. Základnou myšlienkou je prísť so 68 špecifickými bodmi (tzv. orientačné body), ktoré existujú na každej tvári – hornej časti brady, vonkajšieho okraja oka, vnútorného okraja každého obočia atď. Potom budeme trénovať stroj tak, aby bolo možné nájsť týchto 68 špecifických bodov na každej tvári. Teraz, keď vieme, že sú to oči a ústa, budeme jednoducho otáčať a meniť rozsah obrazu tak, aby oči a ústa boli sústredené čo najlepšie viz obr. 2.9. Na to budeme používať len základnú transformáciu obrazu, ako je rotácia a rozsah, ktorý zachováva paralelné linky (tzv. Afinné transformácie).



Obr. 2.9: Transformácia tváre tak, aby bola dokonale centrovaná [18].

Kódovanie tváří. Najjednoduchším prístupom rozpoznávania tváří je priamo porovnať neznámu tvár, so všetkými obrázkami známych ľudí [18]. Keď nájdeme predtým označenú tvár, ktorá vyzerá veľmi podobne ako naša neznáma tvár, musí to byť tá istá osoba.

V skutočnosti je s týmto prístupom obrovský problém. Webové stránky ako Facebook s miliardami užívateľov a milión fotografií nemôžu porovnávať každú označenú tvár s každým novým nahraným obrázkom. To by trvalo príliš dlho. Musia byť schopní rozpoznať tváre v milisekundách, nie v hodinách.

Na rozpoznávanie tváří stačí získať niekoľko základných meraní z každej tváre. Potom sa môže neznáma tvár porovnať so základnými meraniami známych osôb a nájsť tvár s najbližšími meraniami. Môžeme napríklad merať veľkosť každého ucha, vzdialenosť medzi očami, dĺžku nosa . . .

Posledný krok je **vyhľadanie osoby z kódovania**. Tento krok je v skutočnosti najjednoduchším krokom v celom procese. Jediné, čo musíme urobiť, je nájsť osobu v databáze známych ľudí, ktorá má najbližšie merania s neznámou osobou. Všetko, čo musíme urobiť, je trénovať klasifikátor, ktorý môže prijať merania z nového testovacieho obrazu a povedať, ktorá známa osoba je najbližšie k hodnotám merania [18]. Spúšťanie tohto klasifikátora trvá niekoľko milisekúnd. Výsledkom klasifikátora je meno osoby.

3 EXPERIMENTÁLNA ČASŤ

3.1 Databáza Labeled Faces in the Wild:

Jedná sa o databázu obrázkov ľudských tvárí, ktorá bola navrhnutá ako pomôcka pri štúdiu problému rozpoznávania tvárí. [19] Všeobecne termín „rozpoznávanie tváre“ môže odkazovať na niekoľko rôznych problémov ako napríklad:

1. vzhľadom na dva obrázky, z ktorých každý obsahuje tvár, rozhodnúť, či dvaja ľudia predstavujú rovnakú osobu;
2. vzhľadom na obraz tváre osoby rozhodnúť, či ide o konkrétneho jednotlivca (to sa robí na základe porovnania tváre s modelom pre daného jednotlivca alebo inými obrázkami jednotlivca);
3. vzhľadom na obrázok tváre rozhodnúť, ktorá osoba medzi skupinou ľudí, je osoba na obrázku.

Databáza, ktorú nazývame Labeled Faces in the Wild (LFW), je určená na riešenie prvého z týchto problémov, aj keď v prípade potreby môže byť použitá aj na riešenie ostatných problémov.

Hlavný cieľ databázy, je poskytnúť veľkú sadu neobmedzených obrázkov tvárí [19]. Neobmedzené obrázky sú také obrázky, ktoré ukazujú veľký rozsah rozdielov viditeľných v každodennom živote. Patria sem variácie v póze, v osvetlení, vo vyjadrovaní, v pozadí, v rase, v etnickom pôvode, vo veku, v pohlaví, v oblečení, v účese, v kvalite fotoaparátu, v sýtosti farieb, v zaostrení a v ďalších parametroch.

Podrobnosti o databáze, súhrnné štatistiky a vlastnosti databázy sú: [19]

- Databáza obsahuje 13 233 cieľových snímok tváre. Niektoré obrázky obsahujú viac ako jednu tvár. Tvár, ktorá obsahuje centrálny pixel obrazu, ktorý je považovaný za definujúcu tvár obrazu, sa považuje za cieľovú plochu. Tváre iné než cieľová plocha by sa mali ignorovať ako „pozadie“.
- Databáza obsahuje 5 749 adresárov rôznych jednotlivcov. Z nich má 1 680 ľudí dva alebo viac obrázkov v databáze. Zvyšných 4 069 ľudí má len jeden obrázok v databáze.
- Obrázky sú dostupné vo formáte JPEG s rozlíšením 250 x 250 pixelov. Väčšina obrázkov je farebná, hoci pár obrázkov je len v odtieňoch šedej.

Databáza používa súbor obrázkov, ktoré sa zhromažďovali v novinových článkoch na internete [19]. Nie je veľa obrázkov, ktoré sa vyskytujú za extrémnych svetelných podmienok alebo je len veľmi málo obrázkov, ktoré sa vyskytujú za nízkych svetelných podmienok. Existuje obmedzený počet bočných pohľadov na tvári a málo zobrazení zhora alebo zdola. Rozsah a rôznorodosť prítomných obrázkov sú veľmi veľké.

3.2 Testovanie s programom XGBoost

XGBoost je skratka pre „Extreme Gradient Boosting“. [20] Poskytuje implementáciu strojov na zvýšenie gradientu, ktorý vytvoril Tianqi Chen spolupráci s ďalšími vývojármi. Tianqi Chen poskytuje stručný a zaujímavý spätný príbeh o vytvorení XGBoostu v „Story and Lessons Behind the Evolution of XGBoost“ [20]. Je to open source softvérová knižnica, ktorá poskytuje rámec na zvýšenie gradientov pre jazyky C++, Java, Python, R a Julia. Podporuje prácu na systémoch Linux, Windows a macOS.

3.2.1 Cieľ a funkcia

Knižnica je zameraná na laserovú výpočetnú rýchlosť a výkonnosť modelu. Ponúka niekoľko pokročilých funkcií [20].

Modelové funkcie

Implementácia modelu podporuje funkcie aplikácií scikit-learn a R, s novými doplnkami ako je regulácia. Podporujú sa tri hlavné formy zvyšovania gradientu: [20]

- gradient boosting algoritmus nazývaný aj gradient boosting machine vrátane rýchlosti testovania,
- stochastický gradient na zvyšovanie vzorkovania v riadku, stĺpci a stĺpci na rozdelenie úrovní,
- regularizovaný gradient „zvyšovanie regulácie L1 a L2“.

Funkcie systému

Knižnica poskytuje systém na použitie v rôznych počítačových prostrediach, a to nielen: [20]

- paralelizácia konštrukcie stromu pomocou všetkých procesorových jadier počas testovania,
- distribuované výpočty pre testovanie veľmi veľkých modelov,
- výpočty mimo jadro pre veľmi veľké množiny údajov, ktoré nezapadajú do pamäte,
- optimalizácia vyrovňavej pamäte dátových štruktúr a algoritmu na čo najlepšie využitie hardvéru.

Funkcie algoritmu

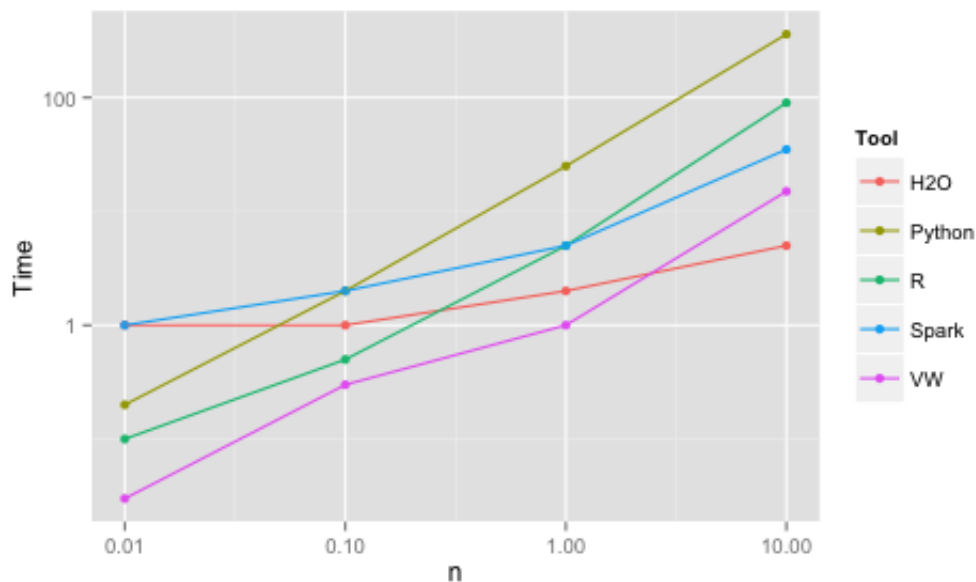
Implementácia algoritmu bola navrhnutá pre čo najlepšiu efektívnosť výpočetného času a pamäťových zdrojov [20].

Niektoré funkcie implementácie kľúčových algoritmov zahŕňajú: [20]

- „roztrúsené vedomie“ – automatické spracovanie chýbajúcich dát,
- „štruktúra bloku“ – podpora paralelizácie konštrukcie stromu,
- pokročilé testovanie – možnosť podporiť už vytvorený model novými dátami.

3.2.2 Výhody XGBoostu

1. Rýchlosť spustenia XGBoostu [20] – všeobecne platí, že XGBoost je v porovnaní s inými implementáciami na zvyšovanie gradientu skutočne rýchly. Szilard Pafka vykonal niekoľko objektívnych porovnávacích kritérií na porovnanie výkonu XGBoostu s inými implementáciami na zvyšovanie gradientu. Svoje výsledky zverejnil na GitHub. Testovanie prebehlo so súbormi s veľkosťami 10K, 100K, 1M, 10M a výsledky sú vidieť na obr. 3.1.



Obr. 3.1: Výsledky testovania rýchlosti[21].

2. Výkonnosť XGBoostu – dominujú štruktúrované alebo tabuľkové súbory dát o klasifikačných a regresných problémoch s predikatívnym modelovaním.

Algoritmus, ktorý XGBoost používa

Knižnica XGBoost implementuje algoritmus zvyšovania gradientu. [20] Tento algoritmus prechádza množstvom rôznych názvov, ako je zvýšenie gradientu, viaceré aditívne regresné stromy, stochastické zvyšovanie gradientu alebo stroje na zvyšovanie gradientu. Zlepšením je technika súborov, ktorá umožňuje pridávať nové

modely na opravu chýb v existujúcich modeloch. Modely sa pridávajú postupne, až kým sa nedosiahne požadované vylepšenie. Zvýšenie gradientu je prístup, pri ktorom sa vytvárajú nové modely, ktoré odstraňujú nedostatky a chyby predchádzajúcich modelov a potom sa pridajú do konečnej predpovede. Nazýva sa to zvyšovanie gradientu, pretože sa používa algoritmus na prechodový gradient, aby minimalizoval stratu pri pridávaní nových modelov. Tento prístup podporuje regresné a klasifikačné predikatívne modelovacie problémy [20].

4 VÝSLEDKY EXPERIMENTÁLNEJ ČASTI

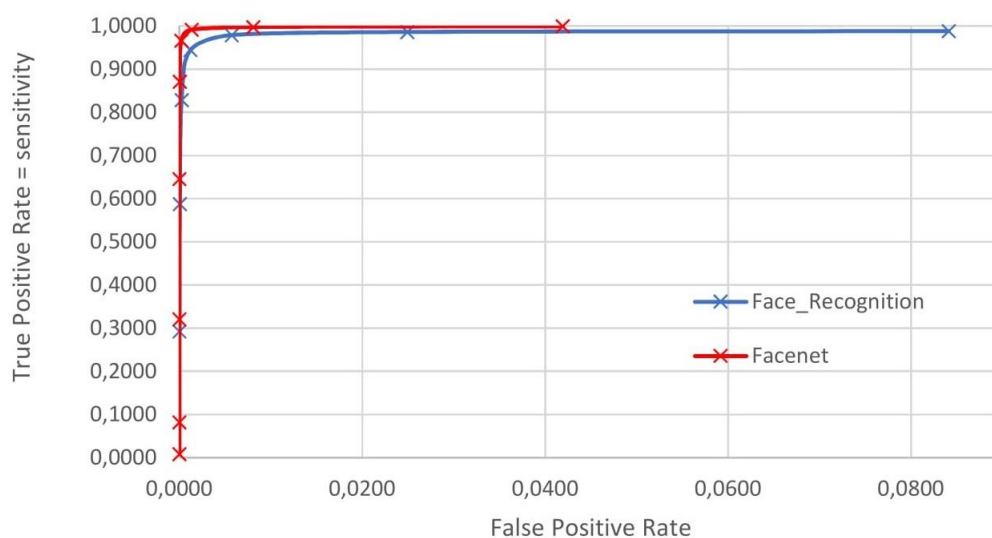
4.1 Výsledky testovania s databázou LFW

Testovanie sa zaoberalo presnosťou a časovou náročnosťou vybraných programov na rozpoznávanie tváre – FaceNet a Face_Recognition. Súbor údajov LFW [19] je štandardným kritériom pri rozpoznávaní tváre. Test sa zaoberá porovnaním dvojice obrázkov a určením či sa jedná o tú istú osobu. Ideálna prahová vzdialenosť obmedzuje nesprávne určenie rozdielnych osôb ako rovnakých – FP (nad prahom) a určenie rovnakých osôb ako rozdielnych – FN (pod prahom). Výsledky testovania presnosti oboch programov sú v tab. 4.1.

Tab. 4.1: Výsledky testovania.

	Vzdialenosť	TP	FP	FN	TN	Total	Specificity	Precision	Accuracy	False Positive Rate	False discovery Rate	True Positive Rate = sensitivity
Face_recognition	0,70	238389	7277653	2912	79225252	86744206	0,9159	0,0317	0,9161	0,0841	0,9683	0,9879
	0,65	237928	2155790	3373	84347115	86744206	0,9751	0,0994	0,9751	0,0249	0,9006	0,9860
	0,60	236183	495783	5118	86007122	86744206	0,9943	0,3227	0,9942	0,0057	0,6773	0,9788
	0,55	227823	104786	13478	86398119	86744206	0,9988	0,6850	0,9986	0,0012	0,3150	0,9441
	0,50	199914	21858	41387	86481047	86744206	0,9997	0,9014	0,9993	0,0003	0,0986	0,8285
	0,45	141703	3641	99598	86499264	86744206	1,0000	0,9749	0,9988	0,0000	0,0251	0,5872
	0,40	70568	596	170733	86502309	86744206	1,0000	0,9916	0,9980	0,0000	0,0084	0,2924
Facenet	1,20	242094	3659679	163	83647592	87549528	0,9581	0,0620	0,9582	0,0419	0,9380	0,9993
	1,10	241648	704332	609	86602939	87549528	0,9919	0,2554	0,9919	0,0081	0,7446	0,9975
	1,00	240202	115495	2055	87191776	87549528	0,9987	0,6753	0,9987	0,0013	0,3247	0,9915
	0,90	233986	18228	8271	87289043	87549528	0,9998	0,9277	0,9997	0,0002	0,0723	0,9659
	0,80	211029	3203	31228	87304068	87549528	1,0000	0,9850	0,9996	0,0000	0,0150	0,8711
	0,70	156342	668	85915	87306603	87549528	1,0000	0,9957	0,9990	0,0000	0,0043	0,6454
	0,60	77656	159	164601	87307112	87549528	1,0000	0,9980	0,9981	0,0000	0,0020	0,3206
	0,50	19887	29	222370	87307242	87549528	1,0000	0,9985	0,9975	0,0000	0,0015	0,0821
	0,40	2051	10	240206	87307261	87549528	1,0000	0,9951	0,9973	0,0000	0,0049	0,0085

Grafickú závislosť medzi nesprávnym určením rozdielnych osôb ako rovnakých a citlivosťou merania je vidieť v grafe 4.1 na základe ROC kriviek. ROC krivka je krivka, ktorá vyjadruje vzťah medzi citlivosťou a špecifitou testu pre rôzne prahové hodnoty. Najlepší výsledok testu sa vyznačuje krivkou ROC, ktorá má najväčšiu plochu pod krivkou. Ideálny test má plochu pod krivkou rovnú 1, čo znamená že má 100% senzitivitu aj špecifitou. Z ROC kriviek v grafe 4.1 vidieť, že lepšie výsledky dosahuje FaceNet. Má vyššiu citlivosť merania pri nižšom počte nesprávne určených rozdielnych osôb ako rovnakých, čo predstavuje chybu testovania pri ideálnom prahu (1) približne 0,13 % čo odpovedá presnosti 99,87 %. Face_Recognition má chybu testovania pri ideálnom prahu (0,55) približne 0,14 %, čo je presnosť 99,86 %.



Obr. 4.1: Závislosť medzi FP a citlivosťou.

Podobné tváre podľa Face_Recognition majú osoby na obrázku 4.2 ale aj mnoho ďalších osôb, ktoré Face_Recognition určil nesprávne ako rovnaké osoby.



Obr. 4.2: Nesprávne určenie rozdielnych osôb ako rovnakých – Face_Recognition.

FaceNet určil že na fotke je tá istá osoba napríklad pri osobách na obrázku 4.3.



Obr. 4.3: Nesprávne určenie rozdielnych osôb ako rovnakých – FaceNet.

Porovnávaním programov z časového hľadiska je rýchlejší FaceNet, pri ktorom trvá výpočet na jednej fotke priemerne 0,146s a pri Face_Recognition trvá výpočet na jednej fotke priemerne 0,454s.

4.2 Výsledky testu s databázou videí a obrázkov

Databázu tvorí 24 neznámych ľudí, z ktorých každý má 75 fotiek. Fotky sú vo veľmi dobrej kvalite a každá osoba má fotky z rôznych uhľov tváre, s bočnými pohľadmi, zavretými očami, miernym úsmevom alebo podsadením brady a vytvorením tzv. dvojitej brady. V databáze má každý človek ľubovoľný počet videí, v obvyklom trvaní 2 až 3 s. Videá sú v horšej kvalite a sú častokrát rozmazané.

Testovanie s databázou obrázkov a videí vo veľmi dobrej kvalite dopadlo horšie ako testovanie s databázou LFW. Výsledky testu sú vidieť v tabuľke 4.2.

Tab. 4.2: Výsledky testovania na databáze s videami

Program	Vzdialenosť	TP	FP	FN	TN	Total	Specificity	Precision	Accuracy	False Positive Rate	False discovery Rate	True Positive Rate = sensitivity	ERROR
FaceNet	1,00	26989	964	2441	675926	706320	0,9986	0,9655	0,9952	0,0014	0,0345	0,9171	0,004821
Face_Recognition	0,55	28776	84	2046	708822	739728	0,9999	0,9971	0,9971	0,0001	0,0029	0,9336	0,002879

Pri testovaní s databázou videí a obrázkov neznámych ľudí sa zhoršila presnosť oboch programov. FaceNet dosahoval presnosť len 99,52 %, čo predstavuje zhoršenie o 0,35 %. Face_Recognition má presnosť 99,71 %, čo znamená že si zhoršil presnosť o 0,15 % oproti testovaniu s databázou LFW.

Výber vzorových obrázkov, ktoré sa nachádzajú v databáze s videami je vidieť na obr. 4.4.



Obr. 4.4: Vzorové obrázky z databázy s videami.

Ďalej testovanie prebiehalo už len s programom Face_Recognition, nakoľko dosahoval lepšie výsledky presnosti a chybovosti pri testovaní s databázou neznámych osôb. Pre zlepšenie výsledkov na testovaní s databázou videí a neznámych osôb bolo vyskúšané vziať do úvahy priemer, medián a smerodajnú odchýlku z vypočítaných príznakov pre každú osobu z databázy neznámych osôb zo 75 fotiek každej osoby. Pre výpočet priemeru zo 75 príznakov pre každú osobu bola použitá funkcia v jazyku python numpy.average. Následne boli výsledky uložené do textového súboru. Podobne bol vypočítaný medián pre každú neznámu osobu zo 75 príznakov. Pre výpočet bola použitá funkcia numpy.median a výsledný vektor bol taktiež uložený do súboru pre každú osobu. Testovanie s databázou videí sa však aj po použití priemeru a mediánu moc nezlepšilo. Výsledky testovania je vidieť v nasledujúcej tabuľke 4.3.

Tab. 4.3: Výsledky testovania s použitím priemeru a mediánu.

	Vzdialenosť	TP	FP	FN	TN	Total	Specificity	Precision	Accuracy	False Positive Rate	False discovery Rate	True Positive Rate = sensitivity	ERROR
PRIEMER	0,55	9692	39	582	236263	246576	0,9998	0,9960	0,9975	0,0002	0,0040	0,9434	0,00251849
MEDIÁN	0,55	9667	32	607	236270	246576	0,9999	0,9967	0,9974	0,0001	0,0033	0,9409	0,00259149

V porovnaní s tabuľkou 4.2 je vidieť, že pri testovaní Face_Recognition s použitím priemeru z vypočítaných príznakov sa výsledná chybovosť znížila o 0,0361 %. Pri testovaní s použitím mediánu sa výsledná chybovosť programu Face_Recognition znížila o 0,0288 %. Tu je vidieť, že vyššie zníženie chybovosti nastalo pri testovaní s použitím priemeru. Chybovosť programu Face_Recognition bolo ďalej skúšané znížiť tým, že sa pridela každej osobe určitá váha. Tomu predchádzal výpočet smerodajnej odchýlky pre každú osobu podľa vzorca 4.1.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (4.1)$$

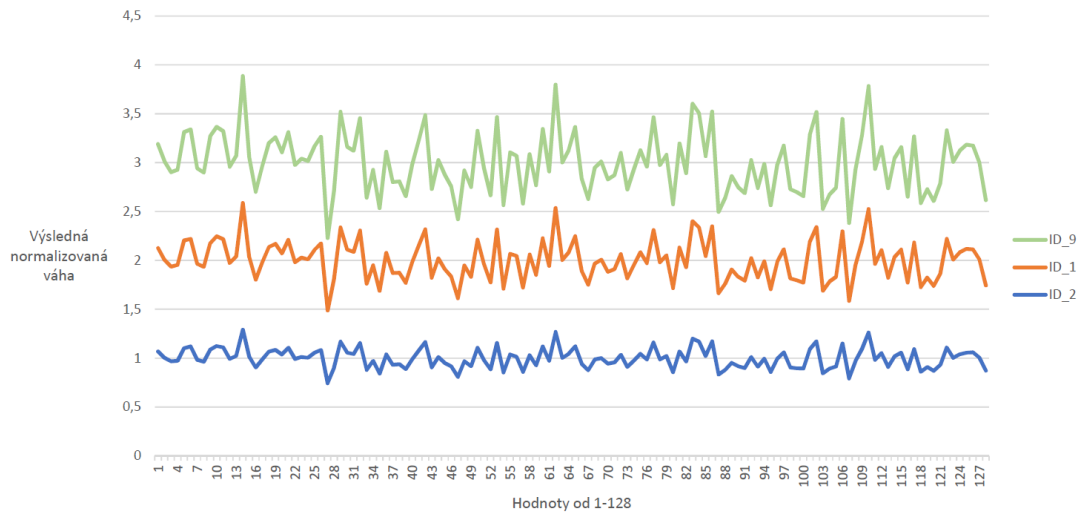
Následne bola vypočítaná aj smerodajnú odchýlku pre databázu LFW a výsledok bol uložený do súboru. Zo smerodajnej odchýlky databázy LFW a zo smerodajných odchýlok neznámych osôb bola následne vypočítaná váha pre každú osobu s použitím vzorca 4.2.

$$W = \frac{\sigma_{LFW}}{\sigma_{osoby} + 1} \quad (4.2)$$

Váhy pre každú osobu boli upravené normalizáciou tak, aby súčet hodnôt v každom vektore váh bol 128. Na normalizáciu bolo potrebné vypočítať konštantu pomocou vzorca 4.3.

$$k = \frac{128}{\sum W} \quad (4.3)$$

Ďalej bola každá hodnota vo vektore váhy vynásobená vypočítanou konštantou každej osoby. Výsledkom bol vektor normalizovanej váhy, ktorého súčet hodnôt bol 128 pre každú osobu. Pre lepšie znázornenie boli normalizované váhy vykreslené pre 3 náhodné osoby. Výsledky normalizovaných váh je vidieť na grafe 4.5. Je zrejmé že, hodnoty normovaných váh sú pre všetky osoby približne rovnaké.



Obr. 4.5: Vykreslenie normovaných váh pre 3 osoby

Normalizované váhy boli použité pri testovaní s databázou videí a výsledky testovania sú zobrazené v nasledujúcej tabuľke 4.4.

Tab. 4.4: Výsledky testovania s použitím priemeru a normovaných váh

	Vzdialenosť	TP	FP	FN	TN	Total	Specificity	Precision	Accuracy	False Positive Rate	False discovery Rate	True Positive Rate = sensitivity	ERROR
PRIEMER S Váhovaním	0,55	9693	49	581	236253	246576	0,9998	0,9950	0,9974	0,0002	0,0050	0,9434	0,00255499

V porovnaní s tabuľkou 4.3 a konkrétnych výsledkov s použitím priemeru je vidieť, že normovány pre každú osobu nepomohli znížiť chybovosť programu Face_Recognition. Naopak, s použitím normováh sa chybovosť zvýšila o 0,00365 % oproti testovaniu iba s priemerom. To môže byť spôsobené tým, že pri teste programu Face_Recognition s použitím len priemerných príznakov každej osoby boli brané do úvahy len fotky neznámych osôb, ktoré sú vo veľmi dobrej kvalite. Pri testovaní s normováhami sa vo výpočte váh zohľadňovala aj smerodajná odchýlka databáze LFW. Nakoľko sú fotky v databáze LFW v zlej kvalite, môže toto byť hlavná príčina, prečo sa pri testovaní s normovanými váhami chybovosť programu Face_Recognition ešte o trochu zvýšila oproti testovaniu s priemerom.

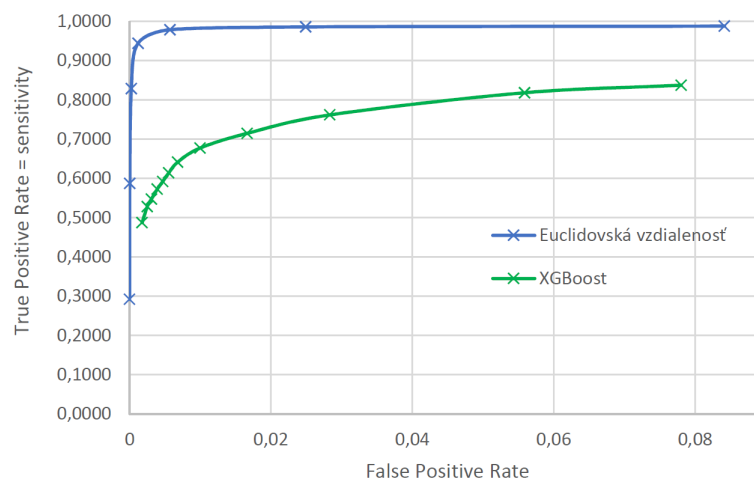
4.3 Výsledky testovania s programom XGBoost

Na automatizáciu porovnávania osôb z databázy LFW s databázou videí a obrázkov neznámych osôb bol použitý program XGBoost. Pre testovanie bol vytvorený pre každú osobu vlastný model, ktorý bol uložený do súboru. Keď bolo vytvorených všetkých 24 modelov, prebehlo načítanie z pamäti a následný výpočet. Výstup XGBoostu boli jednotlivé počty TP, TN, FP a FN z ktorých boli následne vypočítané potrebné štatistické údaje ako napríklad presnosť, chybovosť, senzitivita, falošne pozitívny pomer a podobne. Na nájdenie ideálneho pomeru presnosti a chybovosti bolo potrebné nájsť ideálnu prahovú vzdialenosť. Testovanie bolo začaté s prahovou vzdialenosťou 0,005 a následne bol prah zvyšovaný až po 0,15. Výsledky boli spracované do tabuľky 4.5.

Tab. 4.5: Výsledky testovania XGBoostu

Vzdialenosť	TP	FP	FN	TN	Total	Specificity	Precision	Accuracy	False Positive Rate	False discovery Rate	True Positive Rate = sensitivity	ERROR
0,15	5009	421	5265	235881	246576	0,9982	0,9225	0,9769	0,0018	0,0775	0,4875	0,02305983
0,01	5429	594	4845	235708	246576	0,9975	0,9014	0,9779	0,0025	0,0986	0,5284	0,02205811
0,08	5625	731	4649	235571	246576	0,9969	0,8850	0,9782	0,0031	0,1150	0,5475	0,02181883
0,06	5886	929	4388	235373	246576	0,9961	0,8637	0,9784	0,0039	0,1363	0,5729	0,02156333
0,05	6080	1105	4194	235197	246576	0,9953	0,8462	0,9785	0,0047	0,1538	0,5918	0,02149033
0,04	6311	1313	3963	234989	246576	0,9944	0,8278	0,9786	0,0056	0,1722	0,6143	0,02139705
0,03	6589	1609	3685	234693	246576	0,9932	0,8037	0,9785	0,0068	0,1963	0,6413	0,02147005
0,02	6958	2354	3316	233948	246576	0,9900	0,7472	0,9770	0,0100	0,2528	0,6772	0,02299494
0,013	7345	3935	2929	232367	246576	0,9833	0,6512	0,9722	0,0167	0,3488	0,7149	0,02783726
0,009	7827	6690	2447	229612	246576	0,9717	0,5392	0,9629	0,0283	0,4608	0,7618	0,03705551
0,006	8407	13199	1867	223103	246576	0,9441	0,3891	0,9389	0,0559	0,6109	0,8183	0,06110084
0,005	8605	18437	1669	217865	246576	0,9220	0,3182	0,9185	0,0780	0,6818	0,8376	0,08154078

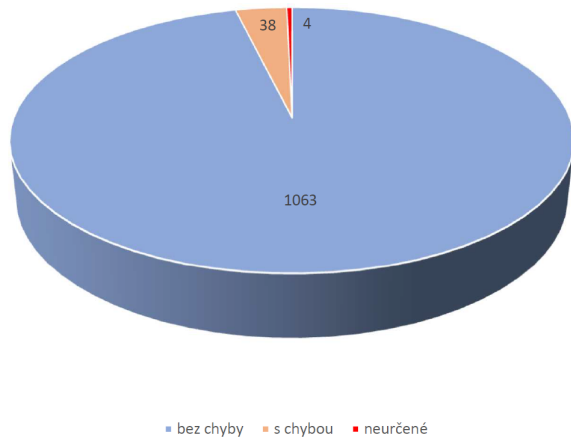
Ako je vidieť v tabuľke 4.5, ideálna prahová vzdialenosť je 0,04. Pri tejto vzdialenosti má program presnosť 97,86 % a chybovosť 2,14 %. Pri zväčšovaní prahovej vzdialenosti sa zhoršovala presnosť aj narastala chybovosť programu, načo má vplyv hlavne fakt, že referenčné fotky neznámych ľudí sú vo veľmi dobrej kvalite oproti testovací videám z tejto databázy. Rozdiel v kvalite obrázkov a videí bolo pokúšané zmierniť rozšírením databázy neznámych osôb o náhodnú úpravu fotiek. Náhodne vybraná fotka bola buď zmenšená na polovicu, Gaussovo rozmazaná alebo bola aplikovaná afinna transformácia. Na všetky tieto úpravy bola použitá priamo funkcia v pythone pomocou knižnice cv2. Vďaka týmto úpravám vznikla väčšia variabilita pôvodných snímok. Avšak, ani napriek zhoršeniu kvality referenčných fotiek sa program nedokázal dobre vysporiadať s rozdielom v kvalite medzi referenčnými obrázkami a testovacími videami. Výsledná štatistika je zobrazená v grafe 4.6.



Obr. 4.6: Závislosť medzi FP pomerom a citlivosťou

Ako vidieť, výsledky testovania s XGBoostom nedosahujú dobré výsledky. V porovnaní s ROC krivkou Euklidovskej vzdialenosti je ROC krivka XGBoostu výrazne horšia. Princiálne bol však postup testovania správny a pri použití iných databáz môžu byť výsledky viditeľne lepšie. V budúcnosti by bolo zaujímavé tieto zistenia rozvinúť a overiť.

Na záver experimentálnej časti bolo otestované rozpoznávanie osôb na jednotlivých snímkoch videí. Databáza obsahuje 1 105 videí. Každá osoba má ľubovoľný počet videí. Testovanie pozostávalo z prechádzania všetkých videí a výpisu osoby, ktorú na snímke program rozpoznal. Za správne určené video bolo považované to, kde najväčší počet rozpoznávaných tvári v jednotlivých snímkoch odpovedal danej osobe. Face_Recognition určil správne 1 063 videí a iba na 38 videách nastala chyba v nesprávnom rozpoznaní osoby na snímke. 4 videá boli tak v zlej kvalite, že program z nich nedokázal rozpoznať žiadnu osobu. Výsledky sú v grafe 4.7.



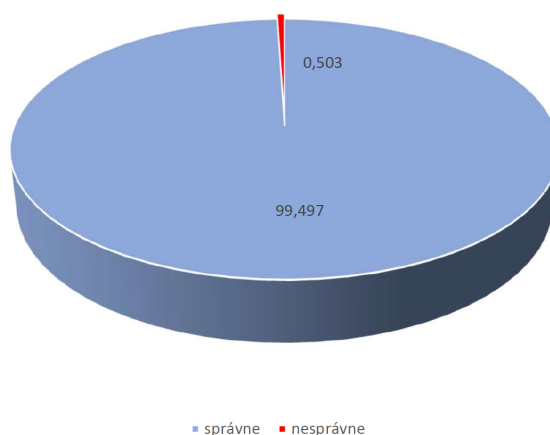
Obr. 4.7: Počet videí určených bez chyby, s chybou a neurčených vôbec

Z celkového hľadiska mal však program dobrú presnosť a primárne určil, že na snímkoch videa prevláda osoba, ktorá v skutočnosti na videu naozaj je. Percentuálne vyjadrenie priemerného počtu správne určených snímok z celkového počtu rozpoznávaných snímok bolo vypočítané podľa vzorca 4.4.

$$x = \frac{TP}{Total} \times 100, \quad (4.4)$$

kde TP je 9 693 (počet správne rozpoznávaných snímok videí) a Total je 9 742 (celkový počet rozpoznávaných snímok vo videách).

Podobne bolo vypočítané aj percentuálne vyjadrenie priemerného počtu nesprávne určených snímok z celkového počtu. Počet nesprávnych snímok je 49. Štatistiku testu je možné pozorovať z grafu 4.8.



Obr. 4.8: Percentuálne vyjadrenie správnych a nesprávnych snímok vo všetkých videách

ZÁVER

Práca sa zaoberala rozpoznávaním ľudských tvárí v obrazoch pomocou algoritmov hlbokého učenia. Cieľom danej práce bolo oboznámenie sa s technikami hlbokého učenia; rozobratie súvisiacich prístupov, najmä konvolučných neurónových sietí; preskúmanie existujúcich programov na rozpoznávanie ľudských tvárí v obraze a vytipovanie dvoch najlepších programov na testovanie z hľadiska presnosti a časovej náročnosti.

Bola uskutočnená inštalácia programov na rozpoznávanie ľudských tvárí – FaceNet a Face_Recognition. Na základe testovania s databázou obrázkov známych ľudí LFW bola stanovená presnosť jednotlivých programov. Tu bolo overené, že lepšie výsledky s databázou LFW dosahuje program FaceNet. Presnosť pri ideálnej prahovej vzdialenosti bola 99,87 %. Presnosť Face_Recognition bola len o trochu nižšia a to 99,86 %, pri ideálnej prahovej vzdialenosti. Pri tejto vzdialenosti program dosahoval chybu testovania len 0,14 %. Pri testovaní programov s reálnymi dátami ľudí vo videu sa presnosť programov trochu znížila. Pri FaceNet bola presnosť na dátovej sade s videami 99,52 % a pri Face_Recognition sa presnosť pohybovala okolo 99,71 %.

Pre zlepšenie výsledkov boli navrhnuté metódy na zníženie chybovosti v prípade, že je k dispozícii viac ako jeden referenčný snímok osoby. Bola otestovaná metóda priemeru a mediánu, pričom lepšie výsledky boli dosiahnuté pri použití priemerného vektoru. Pri tomto teste dosahoval Face_Recognition presnosť 99,75 %. Pri testovaní s použitím mediánu bola presnosť programu 99,74 %. V práci bola overená aj metóda pridelenia váhy každému modelu. Táto metóda sa moc neosvedčila nakoľko Face_Recognition pri nej dosahoval presnosť 99,74 %. Z otestovaných metód je možné považovať za najlepšiu metódu testovanie s priemerom.

Pri teste časovej náročnosti dosahoval lepšie výsledky FaceNet, ktorý ukončil výpočet parametrov na jednej fotke za približne 146 ms, pričom pri Face_Recognition to bolo 453 ms.

Na automatizáciu rozpoznávania osôb bol použitý program XGBoost. Pri ideálnej prahovej vzdialenosti 0,04 dosahoval presnosť 97,86 %. Tento výsledok je výrazne horší ako napríklad pri použití priemerného vektoru. Program sa nedokázal adaptovať na výrazný rozdiel v kvalite referenčných fotiek oproti testovacím fotkám. Na záver experimentálnej časti bolo otestované rozpoznávanie osôb na jednotlivých snímkoch videí. Percentuálny počet správne určených snímkov bol 99,497 % oproti 0,503 % nesprávne určených snímkov.

V práci boli experimentálne otestované presnosti a časové náročnosti programov FaceNet a Face_Recognition. Oba programy dosahovali nižšiu presnosť pri použití databázy s videami, kde bolo k dispozícii viac referenčných snímkov danej osoby. Z výsledkov je najideálnejšie testovanie s použitím priemerného vektoru vypočítaných

príznakov pre dané osoby. Vzhľadom na veľký rozdiel v kvalite referenčných a testovacích obrázkov, testovanie s programom XGBoost dopadlo horšie. Pre zlepšenie výsledkov s programom XGBoost by bolo dobré do budúcnosti overiť presnosť na databázach s menším rozdielom v kvalite obrázkov.

LITERATÚRA

- [1] JAFRI, Rabia; ARABNIA, Hamid R. *A Survey of Face Recognition Techniques*. [online]. Journal of Information Processing Systems, 2009, 41-68 [cit. 2017-11-20]. DOI: 10.3745/JIPS.2009.5.2.041. Dostupné z URL: <https://pdfs.semanticscholar.org/147f/e6bfc76f30ccacc3620662511e452bc395f6.pdf>
- [2] LI, Lufan; JUN, Zhang; FEI, Jiawei; LI, Shuohao. *An incremental face recognition system based on deep learning*. [online]. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), IEEE, 2017, s. 238-241 [cit. 2017-10-19]. DOI: 10.23919/MVA.2017.7986845. ISBN 978-4-9011-2216-0. Dostupné z URL: <http://ieeexplore.ieee.org/document/7986845/>
- [3] BROWNLEE, Jason. *What is Deep Learning?* [online]. 2016, August 16, 2016 [cit. 2018-05-27]. Dostupné z URL: <https://machinelearningmastery.com/what-is-deep-learning/>
- [4] JUNG, Heechul; LEE, Sihaeng; PARK, Sunjeong; KIM, Byungju; KIM, Junmo; LEE, Injae; AHN, Chunghyun; *Development of deep learning-based facial expression recognition system*. [online]. In: Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on, Mokpo, South Korea: IEEE, 1501, s. 1-4 [cit. 2017-10-19]. DOI: 10.1109/FCV.2015.7103729. ISBN 978-1-4799-1720-4. ISSN 978-1-4799-1720-4. Dostupné z URL: <http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/document/7103729/>
- [5] PUMPERLA, Max. *Introduction to Deep Neural Networks*. [online]. [cit. 2017-12-07]. Dostupné z URL: <https://deeplearning4j.org/neuralnet-overview>
- [6] DESHPANDE, Adit. *A Beginner's Guide To Understanding Convolutional Neural Networks*. [online]. [cit. 2017-12-07]. Dostupné z URL: <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>
- [7] URBAN, František. *Človek vs. konvolučné neurónové siete*. Touchit[online]. 2016, 17. JÚLA 2016 [cit. 2018-05-27]. Dostupné z URL: <https://touchit.sk/clovek-vs-konvolucne-neuronove-siete/58956>

- [8] KARN, Ujjwal. *A Quick Introduction to Neural Networks*. [online]. 2016, August 9, 2016 [cit. 2018-05-27]. Dostupné z URL:
<https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/>
- [9] XIE, Danfeng; ZHANG, Lei; BAI, Li. *Deep Learning in Visual Computing and Signal Processing*. [online]. In: Applied Computational Intelligence and Soft Computing, 2017, s. 1-13 [cit. 2017-10-19]. DOI: 10.1155/2017/1320780. ISSN 1687-9724. Dostupné z URL:
<https://www.hindawi.com/journals/acisc/2017/1320780/>
- [10] BALTRUŠAITIS, Tadas; ROBINSON, Peter; MORENCY, Louis-Philippe. *OpenFace: an open source facial behavior analysis toolkit*. [online]. In: Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEE, 2016, s. 1-10 [cit. 2017-10-19]. Dostupné z URL:
<https://www.cl.cam.ac.uk/research/rainbow/projects/openface/>
- [11] AMOS, Brandon; LUDWICZUK, Bartosz; SATYANARAYANAN, Mahadev. *OpenFace: A general-purpose face recognition library with mobile applications*. [online]. In: Carnegie Mellon University Pittsburgh, PA 15213: CMU School of Computer Science, 2016 [cit. 2017-10-22]. ISBN CMU-CS-16-118 Dostupné z URL:
<http://elijah.cs.cmu.edu/DOCS/CMU-CS-16-118.pdf>
- [12] AMOS, Brandon. *OpenFace*. [online]. [cit. 2017-11-27]. Dostupné z URL:
<https://cmusatyalab.github.io/openface/models-and-accuracies/>
- [13] SCHROFF, Florian; KALENICHENKO, Dmitry; PHILBIN, James. *FaceNet: A unified embedding for face recognition and clustering*. [online]. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), IEEE, 2015, s. 815-823 [cit. 2017-10-19]. DOI: 10.1109/CVPR.2015.7298682. ISBN 978-1-4673-6964-0. <http://ieeexplore.ieee.org/document/7298682/>
- [14] SANDBERG, David. *Facenet*. [online]. [cit. 2017-11-27]. Dostupné z URL:
<https://github.com/davidsandberg/facenet/wiki/Validate-on-lfw>
- [15] GRITTI, Tommaso. *Deep learning neural networks and AI explained*. [online]. In: Data & Analytics, 2017 [cit. 2017-11-11]. Dostupné z URL:
<https://www.slideshare.net/braincreators/deep-learning-neural-networks-and-ai-explained>
- [16] GEITGEY, Adam. *Face Recognition*. [online]. [cit. 2017-11-27]. Dostupné z URL:
https://github.com/ageitgey/face_recognition

- [17] *Face Recognition – Dlib*. [online]. [cit. 2017-11-27]. Dostupné z URL:
`http://dlib.net/face_recognition.py.html`
- [18] GEITGEY, Adam. *Machine Learning is Fun! Part 4: Modern Face Recognition with Deep Learning*. [online]. [cit. 2017-12-07]. Dostupné z URL:
`https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition`
- [19] GARY, B. Huang; RAMESH, Manu; BERG, Tamara; LEARNED-MILLER, Erik. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. [online]. In: University of Massachusetts, Amherst, 2007, s. 07-49 [cit. 2017-11-19]. Dostupné z URL:
`http://vis-www.cs.umass.edu/lfw/lfw.pdf`
- [20] BROWNLEE, Jason. *A Gentle Introduction to XGBoost for Applied Machine Learning*. [online]. In: Australia, 2016, August 17, 2016 [cit. 2018-04-15]. Dostupné z URL:
`https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning`
- [21] PAFKA, Szilard. *Simple/limited/incomplete benchmark for scalability, speed and accuracy of machine learning libraries for classification*. [online]. [cit. 2018-04-15]. Dostupné z URL:
`https://github.com/szilard/benchm-ml`

ZOZNAM PRÍLOH

A Obsah priloženého CD

42

A OBSAH PRILOŽENÉHO CD

Na priloženom CD sa nachádza `BakalárskaPráca_VeronikaHornáková.pdf`, ktorá obsahuje textovú časť záverečnej práce. Ďalej CD obsahuje dve ukážkové databázy – `Test_image_database_fromLFW` a `database2`. `Test_image_database_fromLFW` obsahuje 3 osoby z databázy LFW. `Database2` by mali tvoriť referenčné fotky neznámych ľudí. Pre demonštračné účely obsahuje `database2` päť osôb z databázy LFW. Pre testovanie je vhodné doplniť do adresára `Test_image_database_fromLFW` celú databázu LFW. Do `database2` je vhodné dať fotky neznámych osôb a pre testovanie s XGBoostom je nutné pridať do `database2` testovacie videá.

Ďalej sa tu nachádzajú štyri skripty. V príkazovom riadku je vhodné najprv spustiť `Calcelate_Distances.py`, pomocou ktorého sa vypočítajú príznaky z obrázkov z databázy `Test_image_database_fromLFW`. Výsledky sa uložia do `symptoms.txt`. Ako nasledujúce treba spustiť `Modify_Images.py`. `Modify_images.py` vypočíta príznaky na obrázkoch z databázy `database2` a uloží ich do textového súboru do zložky každej osoby. Vypočíta príznaky z pôvodnej fotky a potom náhodne fotku rozmazá, zmenší alebo transformuje a vypočíta príznaky aj z modifikovaných fotiek. Ako nasledujúce sa spustí skript `XGBoost_CreateModel.py`. Tento skript nám vytvorí vlastný model pre každú osobu z `database2` a uloží do textového súboru. Nakoniec pomocou `XGBoost_ProcessModel.py` prebehne načítanie vytvorených modelov a porovnávanie osôb. Program nakoniec vypíše počet TP, TN, FP, FN a total.

```
/ ..... koreňový adresár priloženého CD
├─ BakalárskaPráca_VeronikaHornáková.pdf ..... Text záverečnej práce
├─ Test_image_database_fromLFW ..... Vzorová databáza LFW
├─ database2 ..... Databáza neznámych osôb
├─ Calculate_Distances.py ..... Skript na výpočet príznakov z databázy LFW
├─ Modify_Images.py ..... Skript na výpočet príznakov z database2
├─ XGBoost_CreateModel.py ..... Skript na vytvorenie vlastných modelov
├─ XGBoost_ProcessModels.py ... Skript na načítanie modelov a rozpoznávanie osôb
```