VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
FACULTY OF INFORMATION TECHNOLOGY

ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

HLUBOKÉ NEURONOVÉ SÍTĚ PRO ROZPOZNÁNÍ TVÁŘÍ VE VIDEU

DEEP LEARNING FOR FACIAL RECOGNITION IN VIDEO

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. TOMÁŠ MIHALČIN**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. MICHAL HRADIŠ, Ph.D.**
SUPERVISOR

BRNO 2018

Zadání diplomové práce/21049/2017/xmihal03

Vysoké učení technické v Brně - Fakulta informačních technologií

Ústav počítačové grafiky a multimédií Akademický rok 2017/2018

Zadání diplomové práce

Řešitel: **Mihalčin Tomáš, Bc.**
Obor: Počítačová grafika a multimédia
Téma: **Hluboké neuronové sítě pro rozpoznání tváří ve videu**
Deep Learning for Facial Recognition in Video
Kategorie: Zpracování obrazu

Pokyny:

1. Prostudujte základy teorie neuronových sítí, konvolučních neuronových sítí a zpětné propagace chyb.
2. Vytvořte si přehled o současných metodách pro rozpoznávání tváří pomocí konvolučních hlubokých neuronových sítí ve video sekvencích.
3. Vyberte konkrétní metody a aplikujte je na úlohu rozpoznání tváří ve video sekvencích.
4. Obstarejte si databázi vhodnou pro experimenty.
5. Implementujte navrženou metodu a proveďte experimenty nad datovou sadou.
6. Porovnejte dosažené výsledky a diskutujte možnosti budoucího vývoje.
7. Vytvořte stručné video prezentující vaši práci, její cíle a výsledky.

Literatura:

- Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
- Taigman et al.: DeepFace: Closing the Gap to Human-Level Performance in Face Verification. CVPR 2014.
- Parkhi, Omkar M., Andrea Vedaldi, and Andrew Zisserman. "Deep face recognition." Proceedings of the British Machine Vision 1.3 (2015): 6.
- Yang, Jiaolong, et al. "Neural aggregation network for video face recognition." arXiv preprint arXiv:1603.05474 (2016).

Podrobné závazné pokyny pro vypracování diplomové práce naleznete na adrese http://www.fit.vutbr.cz/info/szz/

Technická zpráva diplomové práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap, které byly vyřešeny v rámci dřívějších projektů (30 až 40% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Hradiš Michal, Ing., Ph.D.**, UPGM FIT VUT
Datum zadání: 1. listopadu 2017
Datum odevzdání: 23. května 2018

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav počítačové grafiky a multimédií
612 66 Brno, Božetěchova 2

doc. Dr. Ing. Jan Černocký
vedoucí ústavu

Abstrakt

Táto diplomová práca sa zameriava na rozpoznanie tvári z videa, konkrétne na spôsob agregácie príznakových vektorov, do jedného diskriminatívneho vektora, tiež nazývaného šablóna. Skúma problém extrémne natočených tvárí, vzhľadom na presnosť verifikácie. Ďalej, porovnáva vzťah medzi šablónami tvorenými vektormi extrahovanými zo snímkov z videa a vektormi z fotografií. Navrhnutá hypotéza je testovaná pomocou dvoch hlbokých konvolučných neurónových sietí a to so známym modelom VGG-16 siete a modelom siete nazývanej Fingera, poskytnutej od firmy Innovatrics. V rámci práce, bolo vykonaných niekoľko experimentov, ktorých výsledky potvrdzujú úspešnosť navrhnutého postupu. Ako metrika presnosti bola zvolená ROC krivka. K práci s neurónovými sieťami bol použitý framework Caffe.

Abstract

This diploma thesis focuses on a face recognition from a video, specifically how to aggregate feature vectors into a single discriminatory vector also called a template. It examines the issue of the extremely angled faces with respect to the accuracy of the verification. Also compares the relationship between templates made from vectors extracted from video frames and vectors from photos. Suggested hypothesis is tested by two deep convolutional neural networks, namely the well-known VGG-16 network model and a model called Fingera provided by company Innovatrics. Several experiments were carried out in the course of the work and the results of which confirm the success of proposed technique. As an accuracy metric was chosen the ROC curve. For work with neural networks was used framework Caffe.

Kľúčové slová

hlboké konvolučné neurónové siete, framework Caffe, rozpoznávanie tvárí, agregácia, konvolúcia, strojové učenie

Keywords

deep convolutional neural network, framework Caffe, face recognition, aggregation, convolution, machine learning

Citácia

MIHALČIN, Tomáš. *Hluboké neuronové sítě pro rozpoznání tváří ve videu.* Brno, 2018. Diplomová práce. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Michal Hradiš, Ph.D.

Hluboké neuronové sítě pro rozpoznání tváří ve videu

Prehlásenie

Prehlasujem, že som túto semestrálnu prácu vypracoval samostatne pod vedením pána Ing. Michala Hradiša, Ph.D. Uviedol som všetky literárne pramene a publikácie, z ktorých som čerpal.

.....
Tomáš Mihalčin
23. mája 2018

Poďakovanie

Rád by som poďakoval Ing. Michalovi Hradišovi Ph.D. za odbornú pomoc, vedenie a ochotu pri tvorbe tejto diplomovej práce a firme Innovatrics za poskytnuté dáta a natrénovanú sieť. V neposledom rade ďakujem virtuálnej organizácií, výpočetnému centru, Metacentrum. Tato práca vznikla s podporou projektov CERIT Scientific Cloud (LM2015085) a CESNET (LM2015042) financovaných z programu MŠMT Projekty veľkých infraštruktúr pre VaVaI.

Obsah

Kapitola 1

Úvod

So zvyšujúcimi sa obavami o bezpečnosť, kamery hrajú dôležitú rolu v spoločnosti a oblasť rozpoznávania tvárí získava čím ďalej, tým väčší význam. Rozpoznávanie tvári, je jednou z aplikácií spracovania obrazu, ktorej sa výskumníci venujú už niekoľko desaťročí 3. Je to biometrický prístup, ktorý poskytuje metódu k rozpoznaniu identity osoby, na základe jej fyziologických vlastností. Rozpoznávanie tvári sa rozšírilo do viacerých oblastí použitia, ako napríklad v biometrických systémoch, v riadení prístupu alebo informačných bezpečnostných systémoch.

V posledných rokoch, sa záujem výskumníkov v oblasti rozpoznávania tvárí presunul z domény statických fotografií, do domény videa. V porovnaní s rozpoznávaním tvárí z fotografií, rozpoznávanie z videa je oveľa náročnejšie. Fotografie v štandardných datasetoch sú zvyčajne zachytené v takmer ideálnych podmienkach alebo dokonca profesionálnymi fotografmi, ako napríklad v datasete Labeled Faces in the Wild(LFW) [19]. V porovnaní s videom, kvalita video snímok zvykne byť výrazne nižšia. Osoby vo videu sú väčšinou v pohybe, čo spôsobuje, že video snímky sú rozmazané, zle zaostrené, v rôznych svetelných podmienkach alebo tvár je v rôznych polohách natočenia. Okrem toho, bezpečnostné kamery alebo mobilné telefóny zvyknú byť menej kvalitné oproti profesionálnym fotoaparátom, čo tiež nepriaznivo vplýva k celkovej kvalite.

Najväčším nedostatkom vo výskume rozpoznávania tvárí z videa je absencia vhodných video datasetov. Na druhú stranu, video obsahuje viac informácií ako fotografia. Video poskytuje viac snímkov osoby, teda je možné zlúčiť niekoľko slabých reprezentácií osoby, do jedného komplexného diskriminačného popisu osoby. Takto vytvorený popis obsahuje mnohokrát väčšie množstvo informácie, ako popis vytvorený len z jednej fotografie. To môžeme vidieť aj vo výsledkoch experimentov 7.

Za posledné roky došlo k ohromnému pokroku v oblasti rozpoznania tváre. Veľkú časť úspechu možno pripísať vývoju techník hlbokého učenia, konkrétnejšie konvolučným neurónovým sieťam (KNN). Tieto techniky stručne vysvetľuje podkapitola 2.5. Zatiaľ čo sa pomocou metód využívajúcich KNN vývoj posunul o markantný krok vpred, ich tréningový proces a využitie vyžaduje, veľké množstvo čistých a správne anotovaných tréningových datasetov a benchmarkov. Verejne dostupným datasetom sa venuje kapitola 4.

Dosiahnuté výsledky niekoľkých najvýkonnejších state-of-the-art algoritmov potvrdili, že sú veľmi blízko [52], ba dokonca prekonali [11, 46], ľudský výkon v oblasti verifikácie tvárí. Tieto postupy a techniky už nasýtili presnosť známej referenčnej metriky Labeled Faces in the Wild (LFW) [19].

Nakoľko LFW benchmark [19] bol predstavený ako mierka presnosti v neobmedzenom prostredí, žiaľ reálnym podmienkam je stále vzdialený. Nie tak dávno, bol navrhnutý nový

benchmark (IJB-A) [27], ktorý si kladie za úlohu sa ešte viac priblížiť k požiadavkám reálneho prostredia. V dôsledku čoho vznikli aj nové problémy a riešenia, ktoré posunuli vývoj rozpoznávania tvárí ešte viac vpred.

Obsah práce je rozdelený do siedmych kapitol. Prvá polovica práce sa venuje úvodu do neurónových sietí 2 a popisuje všeobecnú techniku rozpoznávania tvárí 3. Navrhnutý postup riešenia, spolu s implementáciou 6, je popísaný v druhej polovici práce. V závere sa nachádzajú vykonané experimenty s výsledkami 7.

Kapitola 2

Neurónové siete

Táto kapitola slúži ako úvod do problematiky neurónových sietí.V podkapitole 2.1 je uvedený stručný historický vývoj, 2.2 popisuje všeobecný model neurónu a porovnáva ho s biologickým neurónom. Podkapitola 2.3 sa venuje architektúre neurónových sietí so stručným popisom základných vrstiev.

Ďalšia podkapitola 2.5 obsahuje popis konvolučných neurónových sietí, ktoré sú využité v táto práci. V tejto časti je tiež vysvetlená operácia konvolúcie, technika učenia backpropagation a technika trénovania finetuning.

2.1 História

Začiatok éry neurocomputingu [1] sa datuje od roku 1943, kedy neurofyziológ Warren McCulloch a matematik Walter Pitts publikovali prácu [38] o tom, ako by neuróny mohli teoreticky fungovať. Prácu biologických neurónov demonštrovali návrhom elektrického obvodu, jednoduchej neurónovej sieti. Sieť bola zložená z tzv. formálnych neurónov na symbolickú logiku, na výroky zložené z elementárnych logických operácií. Tento model ešte nebol adaptívny, totiž nebol schopný učenia a jeho váhy boli fixne nastavené pre vykonávanie určitej Boolovskej funkcie. Práca ukázala, že aj jednoduché typy neurónových sietí by v podstate mali byť schopné vypočítať akúkoľvek aritmetickú alebo logickú funkciu. Tento článok bol široko preslávený a mal obrovský vplyv na ďalší vývoj sietí.

V 1949, Donald Hebb napísal knihu *TheOrganizationofBehaivor* [2], v ktorej poukázal na to, že neurónové dráhy sú posilnené každým použitím, čo je vlastne najzákladnejší spôsob, ktorým sa človek učí. Tvrdil, že ak sa dva neuróny súčasne aktivujú, tak sa prepojenie medzi nimi zosilňuje.

Ako sa počítače časom vyvíjali, v 50-tých rokoch 20 storočia, bolo konečne možné simulovať hypotetické neurónové siete. Prvý krok k tomu učinil Nathanial Rochester [37] z výskumných laboratórií IBM. Bohužiaľ, však tento pokus nebol úspešný.

Prvý úspešný neuro-počítač(Mark 1 perceptron) bol vyvinutý medzi rokmi 1957 a 1958 Frankom Rosenblattom, Charlesom Wightmanom a kol. [15]. Pána Rosenblatta dnes poznáme ako zakladateľa Neurocomputingu. Jeho primárny záujem bol, v oblasti rozoznávania vzorov. Ukázal, že McCullochove-Pittsove siete [38] s modifikovateľnými synaptickými váhami je možné natrénovať, tak aby boli schopné rozpoznávať a klasifikovať objekty. To nazval Perceptron.

[1] Umelá neurónová sieť, matematický model navrhnutý tak, aby napodobnil funkciu živých nervových buniek

O niečo neskôr v roku 1959, Bernard Widrow a Marcian Hoff [15] zo Standfordu vyvinuli modely nazývané *ADALINE* a *MADALINE*. Mená vznikli z ich použitia a to viacnásobné adaptívno lineárne prvky (Multiple ADAptive LINear Elements). ADALINE bol vyvinutý na rozpoznanie binárnych vzorov, čiže ak čítal tok bitov z telefónnej linky, dokázal predpovedať nasledujúci bit. MADALINE bola prvá neurónová sieť aplikovaná k riešeniu problémov reálneho sveta. Používala adaptívny filter, ktorý eliminoval ozvenu v telefónnej linke.

Na to v roku 1969, Minsky a Papert vo svojej *knihePerceptrons* [39] ukázali na nedostatok perceptrónov. Dokázali, že tieto siete nie sú vôbec výpočtovo univerzálne a nedokážu riešiť všetky triedy problémov. Išlo hlavne o neschopnosť riešiť lineárne neseparovateľné problémy, ktorým je napríklad logická funkcia XOR 2.2. Tým počiatočné nadšenie z neurónových sietí postupne opadalo a zanechalo za sebou dojem, že výskum neurónových sietí bol len slepou uličkou.

V roku 1986, David Rumelhart, Hinton a Williams [14] prišli s metódou učenia siete, ktorú dnes poznáme pod názvom spätné šírenie chýb. Znamenalo to veľký objav v učení neurónových sietí. V niektorých sieťach dosahovala oproti predošlým metódam niekoľko násobné zrýchlenie procesu učenia. To umožnilo využiť neurónové siete k riešeniu problémov, ktoré do tej doby boli neriešiteľné. Viac o histórií počiatkov neurónových sietí je možné nájsť v [6]

Prvú konvolučnú sieť navrhol japonský výskumník K. Fukushima a kol. v roku 1980 [16]. Táto sieť navrhol k rozoznávaniu ručne písaných číslic.

V roku 1998 LeCun a kol. publikovali prácu [31], v ktorej okrem iného, detailne popísali konvolučné neurónové siete, diskriminačné tréningové metódy, extrakciu príznakov a klasifikáciu. Popísané prístupy demonštrovali aplikovaním rozpoznávania ručne písaných číslic a rozpoznávanie tváre.

Veľký záujem vzbudili konvolučné neurónové siete roku 2012, po prezentovaní metódy Alexanderom Krizhevskym a kol. [28] na súťaži ImageNet Large Scale Visual Recognition Challenge [2]. Krizhevsky a kol. zvíťazili s chybovosťou 15.3 %, čo bolo o 10.9 % menej ako metóda, ktorá obsadila druhé miesto.

Neuronové siete nie sú modely ľudského mozgu

> "Jediný neurón v ľudskom mozgu je neuveriteľne zložitý mechanizmus, ktorému ani vedci dodnes nerozumejú. Neurón v neurónovej sieti je neuveriteľne jednoduchá matematická funkcia, ktorá zachytáva nepatrný zlomok zložitosti biologického neurónu. Takže povedať, že neurónová sieť napodobňuje mozog, je pravda na úrovni voľnej inšpirácie, ale v skutočnosti umelé neurónové siete nemajú veľa spoločného s tým ako biologický mozog funguje.- -Andrew Ng [3]

Ďalším veľkým rozdielom medzi biologickým mozgom a neurónovou sieťou je veľkosť a organizácia. Ľudský mozog pozostáva z omnoho viac neurónov a synapsií než neurónová sieť. Mozog je tiež samo-organizovaný a adaptívny. Neurónové siete, pre porovnanie, sú organizované podľa nejakej architektúry.

Predstavme si to napríklad tak, že neurónová sieť je inšpirovaná biologickým mozgom podobne ako je olympijský štadión v Pekingu inšpirovaný vtáčím hniezdom. To nezna-

[2] Vyhodnotenie algoritmov detekcie objektov a klasifikácie vo veľkom merítku

[3] Standfordský profesor, bývalí líder "Google Brain", zakladateľ vzdelávacieho portálu Coursera a súčasný vedúci výskumnej skupiny v čínskej spoločnosti Baidu

mená, že olympijský štadión je vtáčím hniezdom, ale, že niektoré prvky vtáčieho hniezda sa nachádzajú v dizajne štadióna [54].

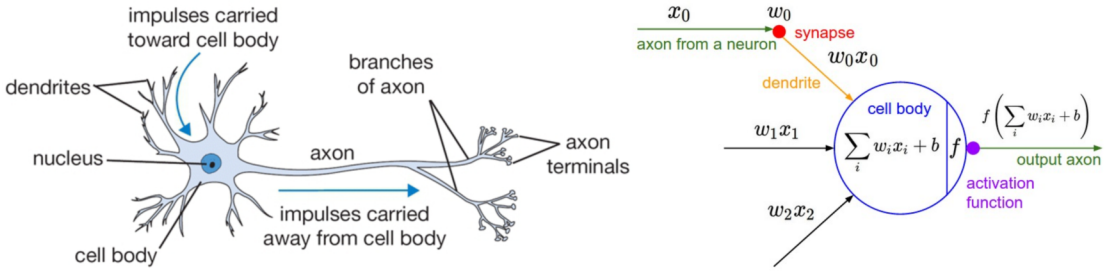
V skutočnosti majú neurónové siete bližšie k štatistickým metódam ako napríklad k nelineárnej regresií alebo regresnej analýze ako k funkcií ľudského mozgu.

2.2 Všeobecný model neurónu

Prv, ako bude popísaný všeobecný model neurónu, je vhodné popísať, na vysokej úrovni abstrakcie, časť biologického systému, ktorou bol umelý neurón z veľkej časti inšpirovaný.

Základná výpočtová jednotka mozgu je neurón. Nervová sústava človeka obsahuje približne 86 miliárd neurónov, ktoré sú prepojené s približne $10^{14} - 10^{15}$ synapsií.

Obrázok 2.1 ilustruje reprezentáciu biologického neurónu na ľavej strane a je všeobecný matematický model umelého neurónu na pravej strane.



Obr. 2.1: Biologický model neurónu, matematický model neurónu. Prebraté z [23].

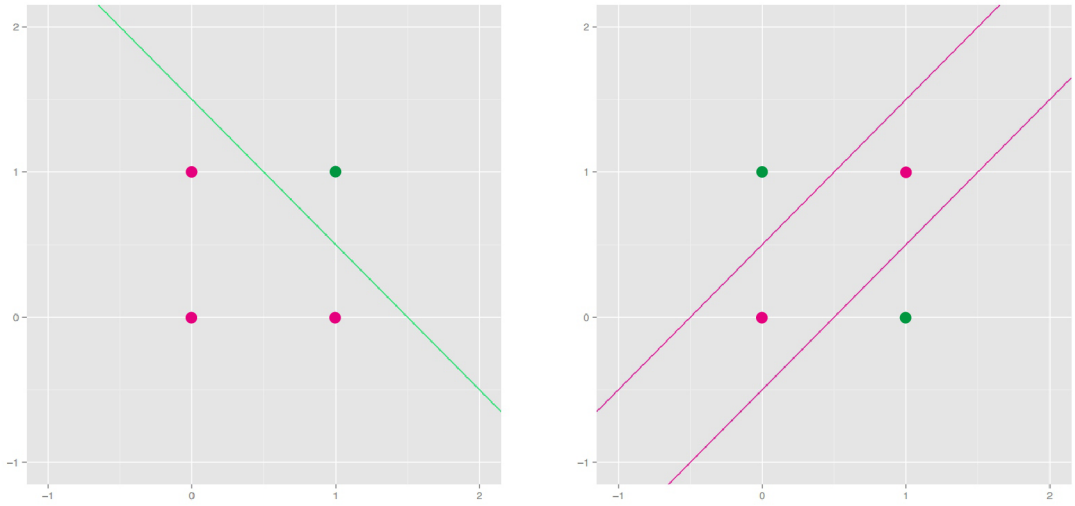
Každý neurón prijíma vstupné signály pomocou svojich vstupných výbežkov (dendridov) a na výstupný výbežok (axon) posiela vytvorený výstupný signál. Výstupný výbežok neurónu, je pripojený cez synapsie k výbežkom ostatných neurónov. Teda synapsie slúžia ku komunikácií medzi jednotlivými neurónmi. V matematickom modely neurónu, signál (x_0), ktorý putuje výstupným výbežkom sa násobí($w_0 x_0$) s výbežkami ďalšieho neurónu na základe synaptickej sily v danej synapsií(w_0). Ideou je, že synaptické sily (váhy w) sú schopné učenia a dokážu ovplyvniť silu a smer, budiče (excitory) (kladná váha) alebo inhibítory (záporná váha), jedného neurónu k druhému. Jednoducho povedané, výbežky nesú signál do tela bunky, v ktorom sú všetky hodnoty sčítané. Táto operácia je znázornená v tele matematického modelu neurónu, v obrázku 2.1 a nazýva sa bázová funkcia. Táto funkcia určuje spôsob výpočtu potenciálu neurónu. Najčastejšie sa používa lineárna bázová funkcia

$$\sum_i w_i x_i. \tag{2.1}$$

Ak je výsledná suma nad nejakú hodnotu prahu, neurón sa aktivuje, tzv. vystrelí, zaslaním výsledného signálu cez jeho výstupný výbežok. Výstup neurónu sa vypočíta z jeho vnútorného potenciálu pomocou aktivačnej funkcie. Jednou z najznámejších aktivačných funkcií je sigmoida 6.1. Vstupom je reálna hodnota (vnútorný potenciál), ktorú prevádza na hodnotu v rozmedzí 0 a 1

$$S(p) = \frac{1}{1 + e^{-p}}. \tag{2.2}$$

Jednoduchý neurón s jedným biasom a skokovou aktivačnou funkciou sa v literatúre nazýva Perceptron [15]. Perceptron je značne limitovaný pretože dokáže klasifikovať množinu len na dve lineárne separovateľné triedy. Na tento vážny nedostatok ukázali aj v roku 1969, Minsky a Papert vo svojej knihe *Perceptrons* [39]. Obrázok 2.2 ilustruje spomínaný problém sepatovateľnosti.



Obr. 2.2: Na ľavej strane sa nachádza logická funkcia AND a na pravej strane logická funkcia XOR.

2.3 Architektúra neurónových sietí

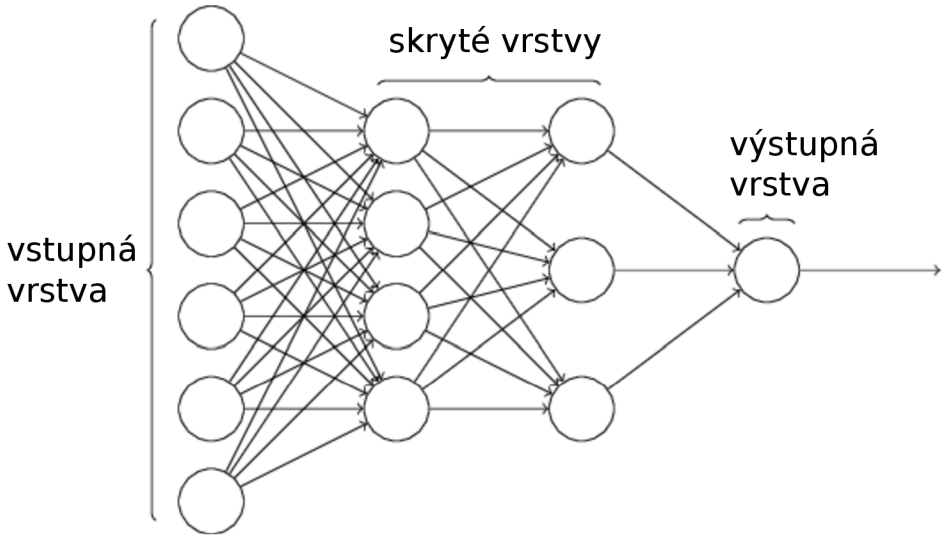
Neurónové siete sú modelované ako kolekcia neurónov, ktoré sú prepojené zvyčajne acyklickým grafom. Modely neurónový sietí sú organizované do určitých vrstiev neurónov [31]. Motiváciou vytvárať neurónové siete je ten, že oproti samotnému neurónu, neurónové siete sú schopné riešiť náročnejšie problémy vrátane klasifikácie lineárne neseparovateľných.

Všeobecný model neurónovej siete 2.3 obsahuje zvyčajne okrem vstupnej a výstupnej vrstvy aj ďalšie vrstvy, ktoré sa nazývajú skryté. Neurónové siete, ktorých výstup z jednej vrstvy je použitý ako vstup ďalšej vrstvy sa volajú dopredné neurónové siete. Avšak existujú aj modely neurónových sietí, ktoré obsahujú slučky. Takéto modely sa nazývajú rekurentné neurónové siete. [34, 41]

V súčasnej dobe, moderné konvolučné siete, obsahujú radovo 100 miliónov parametrov a sú tvorené približne 10-20 vrstvami (odkiaľ pochádza názov hlboké siete, alebo hlboké učenie) [41].

Hlboké neurónové siete pozostávajú z viacerých skrytých vrstiev. Tieto vrstvy sú schopné extrahovať omnoho komplexnejšie príznaky zo vstupných dát.

V sekcií 2.3 bola popísaná len najjednoduchšia architektúra neurónovej site. Existujú mnoho rôznych architektúr neurónových sietí a výkonnosť jednotlivých neurónových sietí je závislá od svojej architektúry a váh. V tejto práci sa zameriavame na konvolučné neurónové siete, ktoré budú popísané v nasledujúcej podkapitole 2.5.



Obr. 2.3: Všeobecná architektúra neurónovej siete. Prebraté z [23].

2.4 Proces učenia

Učením neurónových sietí sa rozumie proces, rovnako ako v Rosenblattovom perceptróne [15], v ktorom sa synaptické váhy menia na základe určitých pravidiel, aby sieť podávala čo najpresnejšie výsledky pre riešený problém.

Proces učenia je rozdelené na kontrolované, nekontrolované učenie a učenie odmeňovaním.

Kontrolované učenie (angl. supervised learning) niekedy tiež nazývané učenie s učiteľom spočíva v tom, že sieť má počas svojho učenia k dispozícií množinu vstupov a k nim referenčné výstupy. V procese učenia sa upravujú váhy tak, aby sa minimalizoval rozdiel medzi výstupom siete a referenčným výstupom.

Nekontrolované učenie (angl. unsupervised learning) , nazývané učenie bez učiteľa alebo samo-organizácia. Sieť má k dispozícií len množinu vstupov a výstup generuje na základe určitých vlastností vstupov za behu. Snaží sa zorganizovať vstupné dáta a objaviť v nich nejaké spoločné vlastnosti. Medzi takéto učenie patrí napríklad Hebbovo učenie [2], ktoré bolo prvýkrát spomenuté v knihe *Organizciasprvania*.

Tretím typom učenia je učenie odmenou (angl. reinforcement learning). Tento prístup je podobný kontrolovanému učeniu v tom, že sieť dostane odozvu do akej miery bola úspešná voči referenčnému výstupu ale s tým rozdielom, že referenčný výstup nemá k dispozícií ale namiesto toho je ohodnotená podľa úspešnosti. Cieľom toho učenia je maximalizácia odmeny, ktorú sieť dostane behom fázy pokus-omyl. Učenie odmenou silno koreluje so správaním zvierat v prírode. Zviera si zapamätá akciu, ktorá jej pomohla dostať potravu, teda odmenu.

Najdôležitejšou metódou k učeniu dopredných hlbokých neurónových sietí je algoritmus spätného šírenia chýb (angl. backpropagation) [30, 41]. Využíva sa pri učení s učiteľom a je založená na zmene hodnôt váh počítaných na základe chyby. Chybou sa myslí rozdiel, medzi očakávaným a skutočným výstupom siete. Hľadá sa teda globálne minimum chybovej

funkcie. Už ako názov hovorí, algoritmus spätného šírenia chýb postupuje v opačnom smere, čiže od výstupnej vrstvy k vstupnej.

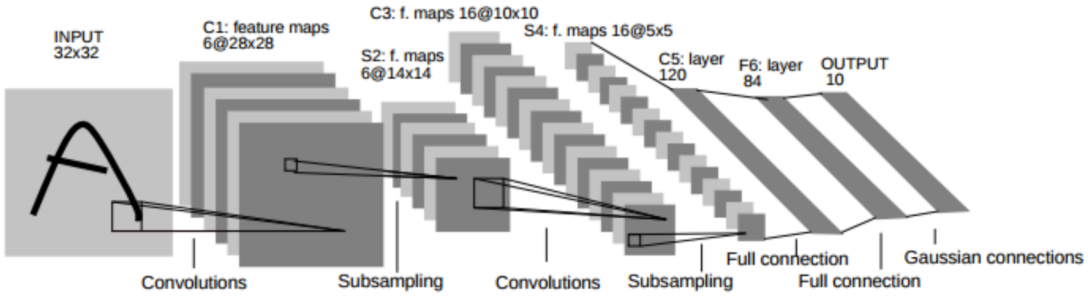
Priebeh učenia je nasledujúci, na začiatku sa všetky hodnoty váh určia náhodne zvyčajne z intervalu $< -0.5, 0.5 >$. Každý prvok z testovacej množiny prejde sieťou a jeho výstup sa porovná s požadovaným výstupom. Vzniknutá chyba sa propaguje naspäť sieťou k predchádzajúcej vrstve. Na základe tejto chyby sú upravené dané váhy. Tento proces sa opakuje dovtedy, kým nie je splnená ukončovacia podmienka, ktorá môže byť napríklad hodnota prahu požadovanej presnosti alebo prekročenie časového limitu.

2.5 Konvolučné neurónové siete

Konvolučné neurónové siete sú špeciálnym druhom dopredných neurónových sietí [24]. Vhodné sú najmä pre štruktúrované dáta ako sú napríklad obraz alebo zvuk. Ako už bolo spomenuté, veľký záujem vzbudili potom, čo s nimi v roku 2012 A. Krizhevsky a kol. [28] zvíťazili na súťaži ImageNet. Používajú sa hlavne v oblastiach problematiky detekcie objektov, klasifikácie objektov, či rozpoznávaní tvárí, ktoré je aj obsahom tejto práce.

Zjednodušene povedané,konvolučné neurónové siete extrahujú význačné vlastnosti zo vstupných dát a transformujú ich na vhodnejšiu reprezentáciu. Konvolučné neurónové siete sa učia pomocou algoritmu spätného šírenia chýb. Vďaka konvolúciám, ktoré sú vykonávané v konvolučných vrstvách, je sieť invariantná voči posunom alebo iným deformáciám vstupného obrazu.

Architektúra konvolučnej siete je zložená z troch typov vrstiev nazývaných konvolučná vrstva, pooling vrstva a plne prepojená vrstva.



Obr. 2.4: Architektúra siete LeNet-5. Konvolučná neurónová sieť navrhnutá k rozpoznávaniu číslic. Prebraté z [31].

Konvolučná vrstva je stavebným kameňom konvolučných neurónových sietí a zároveň prvou vrstvou v sieti. Skladá sa z niekoľkých príznakových máp (angl. feature map), tiež nazývaných aktivačné mapy.[24] Každý neurón v tejto vrstve je prepojený s malým okolím (recepčné pole) z predchádzajúcej vrstvy. Na základe [31] sú neuróny vo vrstve organizované do rovín, v ktorých všetky neuróny zdieľajú rovnakú množinu váh. Množina výsledných neurónov tvorí príznakovú mapu. Všetky neuróny v príznakovej mape počítajú rovnakú operáciu, len na iných miestach obrázku. Zdieľanie váh znižuje pamäťovú náročnosť a umožňuje obmedziť celkový počet výpočtov pri učení a tým pádom zrýchliť proces učenia. Operácia,

ktorú tieto neuróny počítajú sa volá konvolúcia

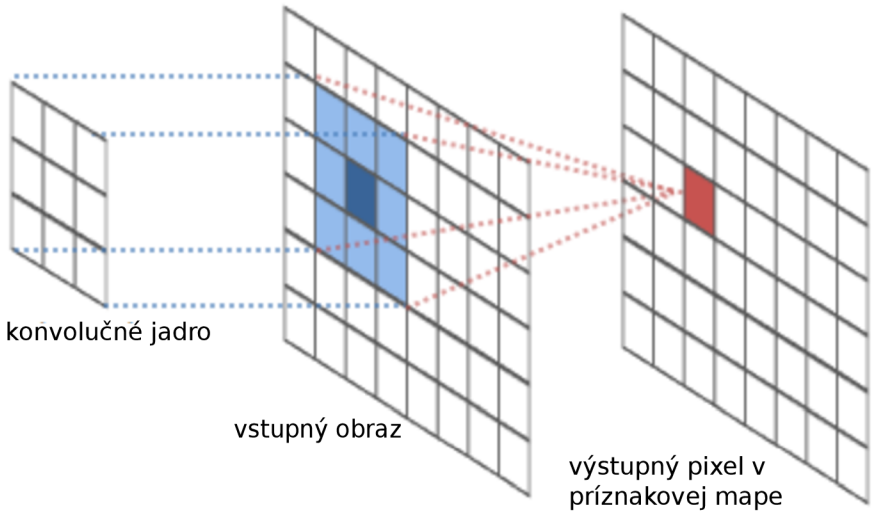
$$I' = I * h. \tag{2.3}$$

Matematicky je diskrétna 2D konvolúcia pre výpočet jedného pixelu na súradniciach x, y, popísaná

$$I'(x,y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x-i, yj)h(i,j) \tag{2.4}$$

, kde symbol $*$ označuje konvolúciu, funkcia $h(x)$ reprezentuje konvolučné jadro [24].

Konvolúcia 2.5 je operácia, ktorá funguje ako filter obrazu. Tento filter, tiež nazývaný kernel alebo konvolučné jadro, je zvyčajne štvorcová maska, ktorá obsahuje váhy neurónovej siete. Výstupný pixel konvolúcie sa spočíta ako súčet hodnôt konvolučného jadra vynásobené s hodnotami vstupného obrazu. Táto operácia sa opakuje dokiaľ konvolučné jadro neprejde celým obrazom.

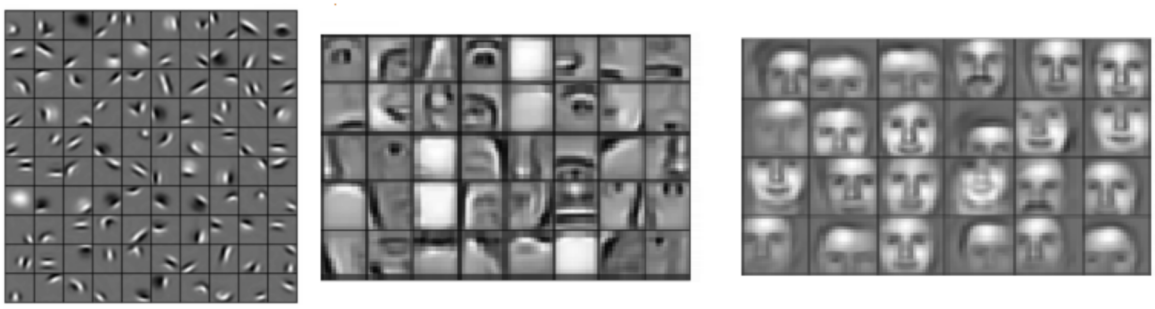


Obr. 2.5: Aplikácia konvolúcie na vstupný obraz. Prebraté z [24].

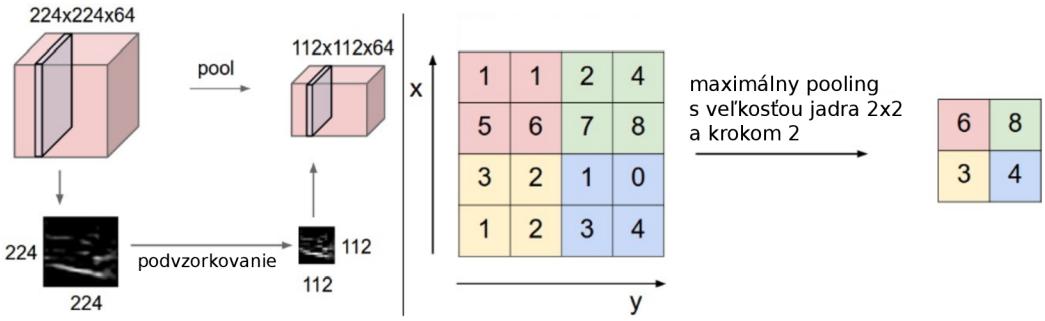
Pomocou neurónov konvolučnej vrstvy je možné zo vstupného obrázku získať základné vlastnosti, ako sú napríklad hrany 2.6 a tiež ich presnú polohu v zdrojovom obraze. Tieto vlastností sú v ďalších vrstvách navzájom kombinované vďaka čomu vznikajú komplexnejšie objekty na vyššej úrovni. Jednotlivé výpočtové konvolučné vrstvy reprezentujú štruktúru obrázku.

V moderných sieťach sa často využívajú viac príznakových máp, čo lineárne závisí aj na dlhšom čase výpočtu. Z toho dôvodu sa bezprostredne za konvolučnými vrstvami umiestňujú pooling vrstvy 2.7. Tieto vrstvy slúžia na tzv. podvzorkovanie príznakových máp. Jedná sa o jednoduchú operáciu kedy sa príznaková mapa rozdelí na neprekrývajúce sa štvorcové oblasti a pixely z každej oblasti sa agregujú do jednej hodnoty. Najčastejšie sa používajú dva typy pooling vrstvy a to maximálny (Max pooling) a priemerný (Average pooling) pooling. Maximálny pooling vyberie maximálnu hodnotu z oblasti hodnôt a tú zapíše na výstup [24]. Krok (angl. stride) určuje, po koľkých pixeloch sa bude podvzorkovávať.

Fine-tuning [21] je metóda, ktorou dokážeme prispôsobiť už natrénovanú sieť k vykonávaniu nami požadovanej funkcie na iných dátach ako bola sieť pôvodne trénovaná. Zvyčajný



Obr. 2.6: Reprezentácie obrazu v jednotlivých vrstvách. Prebraté z [13].



Obr. 2.7: Vľavo: vstupný obraz s veľkosťou [224x224x64] je podvzorkovaný s filtrom o veľkosti 2, krokom(stride) 2 na výstupný obraz o veľkosti [112x112x64]. Všimnite si, že hĺbka ostala zachovaná. Napravo: najznámejšia operácia podvzorkovania max, s krokom 2. Prebraté z [27].

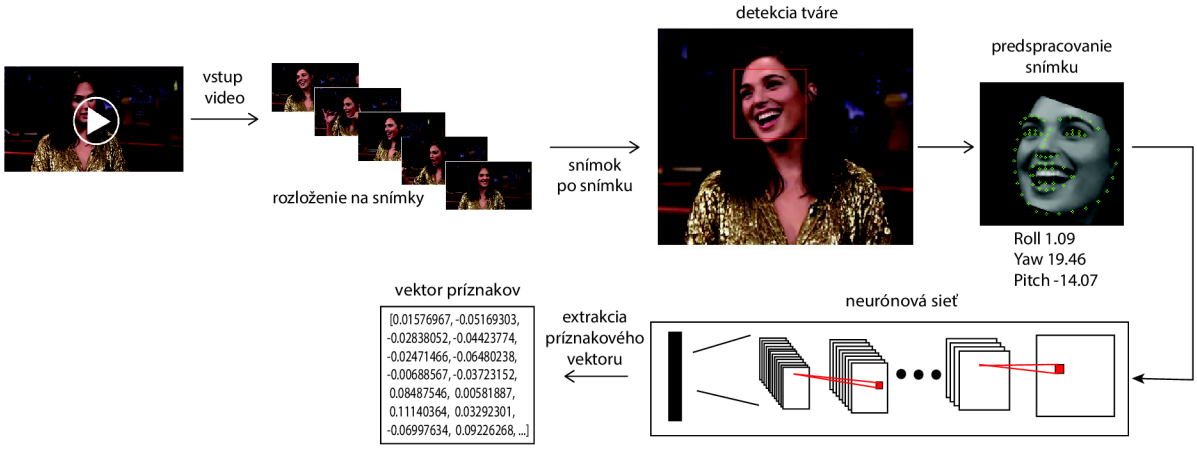
spôsob prispôsobenia už netrénovanej siete, sa vykoná tzv. zmrazením všetkých vrstiev okrem posledných, ktoré dotrénujeme na našich dátach. Teda sieť nemusíme trénovať od začiatku.

Kapitola 3

Rozpoznávanie tvári z videa

Rozpoznávanie tvárí je biometrická metóda, ktorá zahŕňa automatizovaný proces, používaný k verifikácií alebo identifikácií identity osoby. Využíva význačné vlastností tváre a charakteristické črty [47].

Proces rozpoznávania tváre je možné rozdeliť do 4 krokov. Prvým krokom je získanie snímku z obrázku alebo videa. Ďalším krokom je detekcia a následná segmentácia tváre. Tretí krok zahŕňa extrakciu príznakového vektoru. Tento vektor popisuje vlastnosti tváre zo snímky. S takto pripraveným vektorom je možné vykonať požadovanú metódu rozpoznávania alebo verifikácie.



Obr. 3.1: Postup metódy rozpoznávania tváre.

3.1 Postup metódy rozpoznávania tváre

Detekcia tváre je už mnohé roky aktívnou oblasťou výskumu počítačového videnia. K detekcií tváre bolo navrhnutých niekoľko prístupov [33, 10, 36, 32].

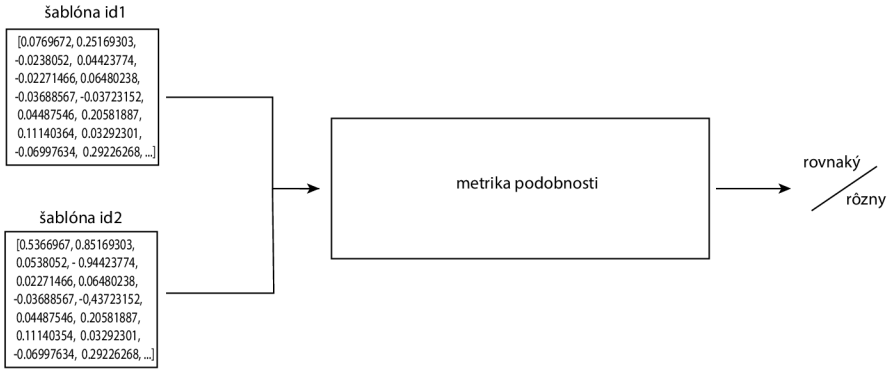
Medzi najznámejšie metódy patria, kaskádová detekčná metóda tvárí predstavená v roku 2001 pánmi Viola a Jones [55], metóda LBP (lokálne binárne vzory) [1], ktorá popisuje vlastnosti obrazu pomocou príznakov, ktoré obraz charakterizujú, AdaBoost algoritmus, čo je meta-algoritmus strojového učenia predstavený Y. Freudom a R. Schapirom [18], ktorý

zaň vyhrali v roku 2003 Gödelovú cenu, SMQT príznaky and SNOW klasifikačná metóda [45] a neurónové siete [43].

Po detekcií a segmentácií tváre, proces extrakcie príznakov zvyčajne predchádza proces predspracovania. Predspracovanie snímkov môže byť vykonané rôznym natočením, odstránením šumu, orezaním alebo zmenšením veľkosti a podobne.

Extrakcia príznakov slúži na vytvorenie novej reprezentácie vstupného obrazu. Danou reprezentáciou je vektor príznakov, ktorý si môžeme predstaviť ako odtlačok prsta, ktorý jednoznačne rozlišuje danú tvár. K extrakcií príznakov sa využívajú konvolučné neurónové siete popísané v 2.5.

Metódy rozpoznávania Verifikácia je proces, v ktorom sa zisťuje či dva vstupné tváre patria tej istej osobe alebo nie. Je to tzv. porovnanie $1 : 1$. Úlohou je zistiť či skúmaná osoba je tá, za ktorú sa vydáva. 3.2



Obr. 3.2: Výpočet vzdialenosti vektorov.

Identifikácia je proces kedy sa vstupná tvár porovnáva so všetkými osobami v databáze. Úlohou je zistiť komu patrí tvár na snímke. Je to problém $1 : N$, kde n je počet osôb v databáze.

Kapitola 4

Dostupné datasety

Táto kapitola stručne popisuje verejne dostupné datasety, ktoré sú vhodné pre metódy rozpoznávania tvárí.

4.1 Postupný vývoj datasetov

Medzi najviac známy bezpochyby patrí dataset Labeled Faces in the Wild (LFW) [19] z roku 2007. Obsahuje 5 749 identít s 13 000 fotografiami. Tento dataset definuje aj spôsob merania presnosti, tzv. benchmark, ktorý sa vyhodnocuje pomocou 10-cross validácie. Ako bolo už spomenuté, rýchlym vývojom neurónových sietí bol tento benchmark nasýtený. To podporilo vznik nových datasetov a benchmarkov inšpirovanými práve datasetom LFW. V tabuľke 4.1, ktorá obsahuje verejne dostupné datasety v chronologickom poradí vidieť, ako postupom času vznikali, čoraz väčšie a náročnejšie datasety.

Obr. 4.1: Ukážka fotografií z LFW datasetu.

Dataset	# subjektov	# fotografií	# videí	# cca. fotografií na subjekt	# cca. videí na subjekt	rok
LFW [19]	5,749	13,233	0	2.3	0	2007
PubFig [29]	200	58,797	0	294	0	2009
YTF [57]	1,595	0	3,425	0	2.1	2011
PaSC [5]	293	9,376	2,802	32	9.6	2013
CASIA-WebFace [**?**]	10, 575	494,414	0	46.7	0	2014
FaceScrub [40]	695	141,130	0	202.75	0	2014
CelebA [62]	10,177	202,599	0	19.9	0	2015
VGGFace [42]	2, 622	982,803 (*2.6M [1])	0	375 (*1000)	0	2015
IJB-A [27]	500	5,712	2,085	11.4	4.2	2015
CACD [8]	2,000	163,446	0	78.4	0	2015
MegaFace [25]	690, 572	1M	0	-	0	2016
WIDER FACE [61]	-	32, 203	0	-	0	2016
MS-Celeb-1M [17]	100,000	10 M	0	100	0	2016
UMDFaces [4]	8,277	367,888	0	43.3	0	2016
IJB-B [56]	1,845	11,754	7,011	6.37	3.8	2017
UMDFaces-Videos [3]	3, 107	0	22,075	0	7.1	2017
VGGFace2 [7]	9,131	3.31 M	0	-	0	2018

Tabuľka 4.1: Zoznam najväčších verejne dostupných datasetov.

VGGFace dataset, publikovaný v roku 2015, obsahoval 2.6 milióna fotografií s 2 622 identitami. Patril medzi najväčšie voľne dostupné datasety. Avšak obsahoval príliš veľa nepresných anotácií, ktoré museli byť časom manuálne odstránené. Vyčistený dataset obsahoval 800 000 fotografií s približne 305 fotografiami k jednej identite.

Prelomom nastal v roku 2016, kedy Microsoft publikoval Ms-Celeb-1M dataset so 100 000 celebritami. Každá identita je zachytená 100 fotografiami, teda dataset obsahuje 10 miliónov fotografií pre trénovanie aj testovanie, a tým sa stáva najväčším verejným datasetom.

Okrem verejne dostupných datasetov, veľké spoločnosti ako Facebook a Google disponujú radovo väčšími internými datasetmi. Napríklad, Facebook [53] v roku 2015 natrénoval model k identifikácií tvárí na datasete veľkom 500 miliónov fotografií s viac ako 10 miliónmi subjektov. Rok predtým, teda v roku 2014, použili na trénovanie hlbokej neurónovej siete 4.4 miliónov fotografií so 4 000 subjektmi [52]. Google tiež použili dataset s cez 200 miliónmi fotografií s viac ako 8 miliónov subjektov k trénovaniu siete so 140 miliónmi parametrov. Z toho vyplýva, že akademická sféra je značne nezvýhodnená.

Na rozdiel od spomínaných datasetov, ktoré sa zameriavajú na rozpoznávanie tvárí zo statických fotografií, Youtube Faces (YTF) [57] a UMDFaces-Videos [3] datasety sú mierené na rozpoznávanie tvárí z videí v reálnych podmienkach. Dataset YTF obsahuje 1 591 subjektov a 3 425 videí, zatiaľ čo UMDFaces-Videos dataset je väčší s 3 107 subjektmi a 22 075 videami. Výhodou UMDFaces-Videos datasetu je, že jeho subjekty sú podmnožinou subjektov z datasetu UMDFaces.

K vypracovaní tejto práce boli zvolené datasety IJBA, UMDFaces a UMDFaces-Videos.

[1] Práca [14] tvrdí, že po konečnom manuálnom filtrovaní, finálny počet vhodných fotografií je 982,803 tvorených približne z 95% čelných a 5% profilových natočení.

4.2 IARPA Janus Benchmark-A

IJB-A dataset [56] bol vytvorený za účelom poskytnúť novší a náročnejší dataset pre verifikáciu a identifikáciu. Obsahuje fotografie a videá subjektov, ktoré boli manuálne anotované s využitím Mechanical Turk (MTurk) [2]. Anotácie obsahujú informáciu o ohraničení tváre, referenčných bodoch tváre, pozíciu očí či nosu. Subjekty boli zachytené v nekontrolovanom prostredí tzv. $in-the-wild$, za účelom priblížiť dataset k čo najviac reálnym podmienkam. Subjekty boli zámerne vybrané aby vytvorili širšie geografické rozloženie než predchádzajúce datasety. Príklady fotografií a videí z datasetu zobrazuje obrázok 4.2.

Obr. 4.2: Príklady tvárí z datasetu IJB-A. Tieto fotografie a video snímky zdôrazňujú mnohé kľúčové charakteristiky verejne dostupného datasetu, zahŕňajúc pózy v plnom rozsahu natočenia, zmes fotografií a videí a širokú škálu rôznych podmienok a subjektov geografického pôvodu. Prebraté z [27].

Dataset obsahuje 500 subjektov s 5 396 fotografiami a 2 042 videami, teda 11.4 fotografií a 4.2 videí pre jeden subjekt. Široká škála natočení tvárí a rôzne svetelné podmienky, výrazy tváre a rozlíšenie sú hlavné atribúty, ktoré robia IJB-A dataset veľmi náročný. Graf 4.3 znázorňuje rozloženie počtu fotografií a videí pre jeden subjekt.

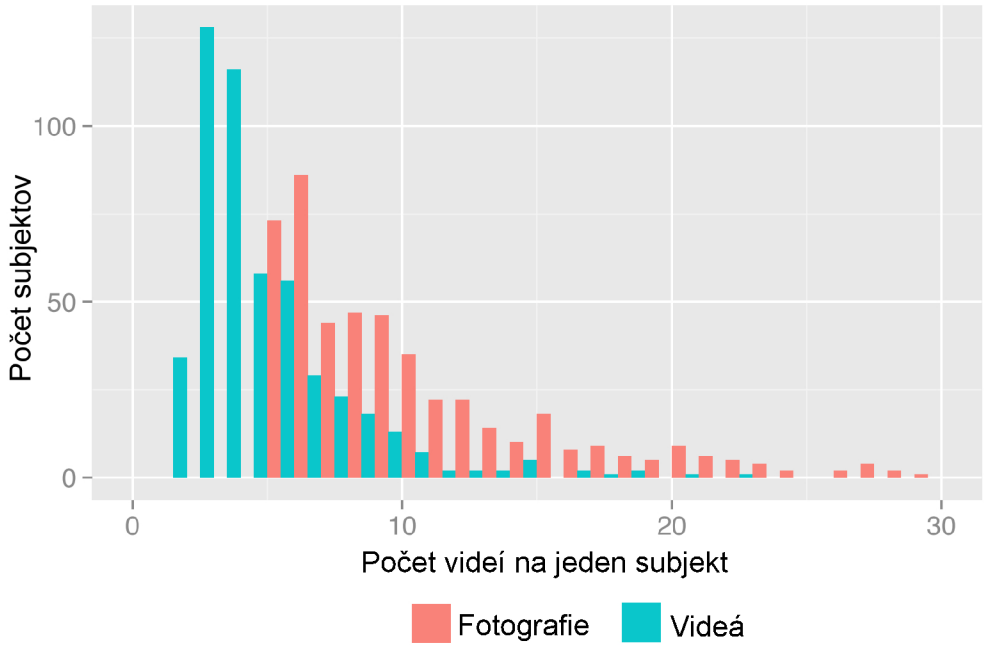
IJB-A poskytuje tiež protokoly pre evaluáciu verifikácie aj identifikácie. Evaluácia je založená na tzv. šablónach, ako najmenšia reprezentačná jednotka, teda namiesto vyhodnocovania štýlom fotografia voči fotografií. Šablóna je množina všetkých médií (fotografií a/alebo videí) subjektu, zkombinovaných do jednej súhrnnej reprezentačnej jednotky 5.1.

Verifikačný protokol je podobný iným benchmarkom [19, 57]. Obsahuje 10 rozdelení. Každé rozdelenie obsahuje približne 11748 párov šablón. Protokolom je predpísaný presný zoznam dvojíc, ktoré pozitívne (porovnáva sa rovnaký subjekt) a negatívne(porovnávajú sa dva rozdielne subjekty) dvojice sa budú porovnávať v ktorom rozdelení. Obrázok 4.4 obsahuje snímky a fotografie z rozdelenia č. 1. Pre lepšie pochopenie náročnosti IJB-A benchmarku obrázok 4.1 obsahuje fotografie z LFW benchmarku. Už na prvý pohľad vidno, že fotografie z LFW datasetu sú kvalitné a natočenie tvárí je takmer frontálne, čo sa o datasete IJB-A povedať nedá.

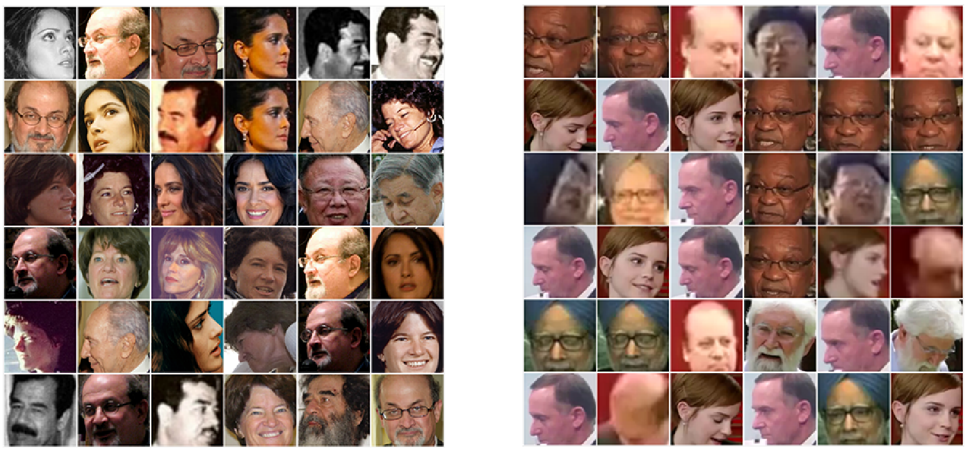
4.3 UMDFaces

Tento dataset patrí rovnako ako IJB-A medzi novšie datasety, ktorý obsahuje videá aj fotografie. Je rozdelený do dvoch častí. Prvá časť obsahuje fotografie 8 277 subjektov s 367 888 anotovanými tvárami. V druhej časti sa nachádzajú 22 000 videí stiahnutých z Youtube s 3 100 subjektmi. Videá sú rozdelené do vyše 3.7 miliónov anotovaných snímkov. Všetky

[2] Crowdsourcing služba poskytovaná firmou Amazon



Obr. 4.3: Rozloženie fotografií a videí v datasetu IJB-A na jeden subjekt. Prebraté z [27].



Obr. 4.4: Ukážka fotografií a snímkov z rozdelenia č.1 z IJB-A benchmarku.

anotácie boli manuálne vytvorené pomocou Mechanic Turks, teda disponuje minimálnymi chybami.

UMDFaces [4] definuje aj nový protokol pre verifikáciu tvárí. Existujú tri stupne náročnosti evaluácií verifikácie a to, ľahký, stredný a ťažký stupeň. Každá evaluácia obsahuje 100 000 presne definovaných párov. Ukážky fotografií v jednotlivých stupňoch náročnosti zobrazuje obrázok 4.5.

Celý dataset je verejne dostupný až na videá. K stiahnutí videí je nutné kontaktovať pána Ankana Bansala z[3], ktorý mi sprístupnil adresy k stiahnutiu celých videí (1.2TB).

[3]Pre sprístupnenie videí k stiahnutiu: ankan@umiacs.umd.edu

Obr. 4.5: Ukážka fotografií z protokolov zľava: ľahká, stredná a ťažká náročnosť

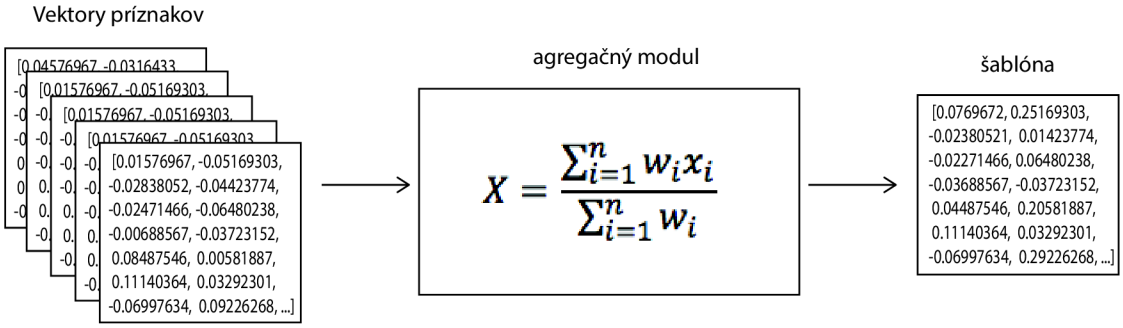
Kapitola 5

Agregácia a existujúce riešenia

V tejto kapitole je vysvetlená idea a základné techniky agregácie. Ďalej sú predstavené existujúce riešenia, ktoré dosahujú state-of-the-art úspešnosť.

5.1 Agregácia

Agregácia je technika, pri ktorej sa z niekoľko vstupných fotografií a/alebo videí vytvára jedna komplexná reprezentácia subjektu. Agregáciu sa využíva napríklad v prípade implementácie systému pre kontrolu prístupu do objektu. Proces agregácie je znázornený na obrázku 5.1. Pred samotnou agregáciou je ešte potrebné vykonať pár krokov, ktoré sú zobrazené na grafe 3.1.



Obr. 5.1: Spôsob agregácie vstupných vektorov.

Nasledujúci príklad stručne vysvetľuje použitie agregácie. Kamerový systém zachytí na video subjekt. Video je rozložené do video snímkov, ktoré zachytávajú daný subjekt v rôznych polohách natočení a svetelných podmienkach. Tieto snímky prídu na vstup konvolučnej neurónovej siete. Táto sieť extrahuje zo snímok vektor príznakov, ktorý je definovaný pre každý snímok zvlášť. Príznakové vektory prídu na vstup agregačnému modulu. Výstupom agregácie je jeden príznakový vektor, ktorý reprezentuje všetky vstupné vektory, teda subjekt z videa. Takto vytvorený reprezentačný vektor sa v literatúre nazýva šablóna (angl. template). Pomocou takto vytvorených šablón už nie je problém vykonať metódy rozpoznávania, verifikáciu alebo identifikáciu.

Spôsob agregácie je predmetom viacerých štúdií. Medzi najjednoduchší spôsob agregácie patrí metóda nazývaná pooling [60]. Ide o techniku zlučovania určitých hodnôt na rovnakých

indexoch z príznakových vektorov. Výber hodnôt zvyčajne býva minimum, maximum alebo priemer. Pooling sa vyznačuje rýchlosťou a nie je náročný na výpočtovú kapacitu.

Jiaolong Yang a kol. [60] publikovali spôsob agregácie s využitím neurónovej siete. Táto metóda je jednou z ďalších možností agregácie. Ich agregačná sieť sa vyznačuje nezávislosťou na počte a poradí vstupných dát. Poukázali na to, že táto metóda dosahuje lepšie výsledky, ako metóda pooling.

Prvý dataset, ktorý definoval protokol verifikácie práve na základe šablón je spomínaný IJB-A [56].

Motiváciu takto vytvárať agregované deskriptory oproti vektoru príznakov extrahovaného z jednej fotografie je práve množstvo obsiahnutej informácie. Video-sekvencia zachytáva subjekt vo viacerých pózach, čo pomáha k vytvoreniu kvalitnejšieho výsledného vektoru.

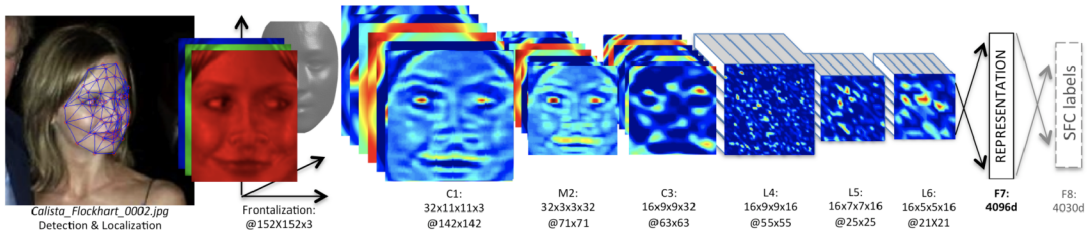
5.2 Existujúce riešenia

Všetky techniky, ktoré dosahujú najlepšie výsledky v oblasti verifikácie tváre na Labeled Faces in the Wild [19] a YouTubeFaces [15] sú založené na konvolučných neurónových sieťach.

VGG-Face využíva architektúru VGG-16 konvolučnej siete [48], ktorá je trénovaná na novom vyčistenom datasete obsahujúcom 2.6M fotografií z 2 622 subjektov. Táto reprezentácia využíva triplet loss embedding, 2D zarovnanie k normalizácií a vykazuje state-of-the-art úspešnosť.

FaceNet [46] aplikoval architektúru Inception [51] k riešení problému verifikácie tváre. V tomto prístupe využili k trénovaniu metriku euklidovského priestoru a triplet loss embedding k vyprodukovaniu 128 dimenzionálnej reprezentácie. Táto sieť bola natrénovaná na súkromnom datasete obsahujúcom cez 200 fotografií.

DeepFace [52, 53] použili hlbokú sieť spojenú s 3D zarovnaním k normalizácií tváre. Pomocou orientačných bodov tváre vytvorili 3D model 5.2. Vďaka tomu každá tvár na vstupe bola transformovaná do frontálnej podoby, kde sa pre každú tvár nachádzali hlavné body tváre zhruba na rovnakom mieste. Tento prístup ukázal zlepšenému výkonu siete.



Obr. 5.2: 3D zarovnanie použité v architektúre DeepFace. Prebraté z [52].

DeepID2+ [50] a DeepID3 [49] rozšírili architektúru Inception zahrnutím Joint Bayesian metriky [9] a multi-task učenia a tým dosiahli ešte lepších výsledkov.

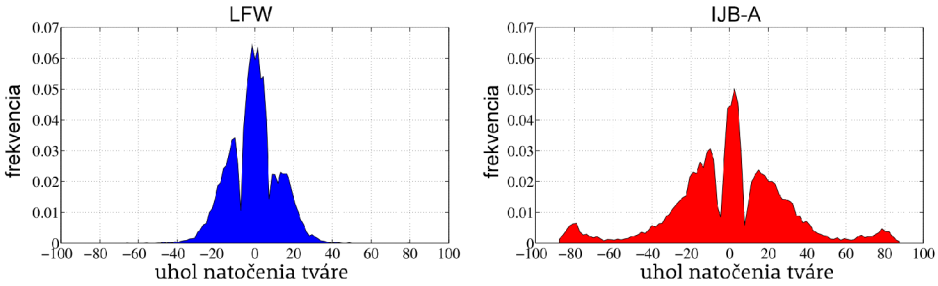
Všetky tieto najvýkonnejšie prístupy zdieľajú spoločné vlastnosti. Všetky využívajú hlboké konvolučné neurónové siete a všetky vykonávajú určitú formu normalizácie pozície tváre ako napríklad 2D či 3D zarovnanie. Štúdovaním toho javu, v tejto práci experimentujem so spôsobom agregácie vzhľadom na natočenie tváre.

Kapitola 6

Návrh a implementácia

Táto kapitola obsahuje návrh agregačného modulu príznakových vektorov, popis použitých neurónových sietí a nástrojov k implementácií. Ako vysvetľuje podkapitola 5.1 existuje viacero techník agregácií a ďalšie postupy sú stále predmetom viacerých štúdií.

Tváre z datasetov IJB-A a UMDFaces obsahujú voči datasetu LFW natočenie v širšom rozsahu, čo dokazujú aj grafy 6.1.



Obr. 6.1: Rozloženie rozsahu natočenia tvárí. Prebraté z [35]

Snímky s extrémne natočenou tvárou neposkytujú také množstvo informácie ako frontálne snímky tvárí 6.2

Obr. 6.2: V prvom rade je ukážka nežiadúcich snímok z datasetu UMDF [3]. V druhom rade požadované snímky.

6.1 Návrh agregačného modulu

Navrhnutá agregácia sa zameriava na kvalitu vstupných snímkov vzhľadom k póze a uhlu natočenia tváre. Ohodnotí každý vstupný snímok váhou, ktorá určí ako veľmi daná snímka

ovplyvní výsledný vektor. Skúma kvalitu výsledných šablón. Výhodou tohto prístupu je eliminácia málo kvalitných snímkov, ktoré neposkytujú dostatočné množstvo informácie. V prípade klasického aritmetického priemeru nie je možnosť vyselektovať snímky s nízkou informačnou hodnotou 6.2, ktoré vstupujú do výpočtu šablóny. Spôsob agregácie sa počíta váženým priemerom

$$x = \frac{\sum_n^{i=1} w_i x_i}{\sum_n^{i=1} w_i}, \tag{6.1}$$

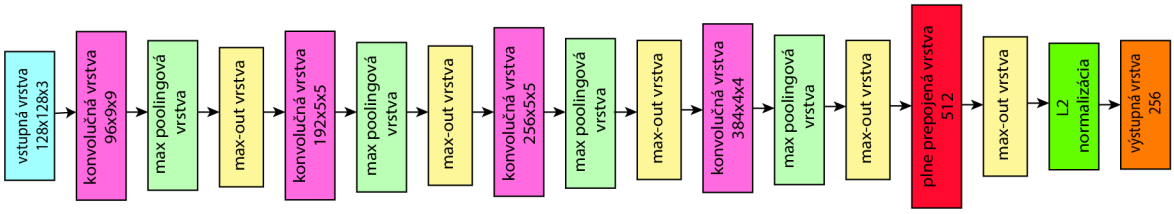
kde x je príznakový vektor snímky a w je jej váha.

V rámci experimentov v kapitole 7 je skúmaná úspešnosť príznakových vektorov agregovaných aritmetickým priemerom voči agregácií ohodnotených vstupných dát.

6.2 Použité neurónové siete

Kvôli väčšej presnosti výsledkov, bola implementácia a všetky experimenty vykonávané pomocou dvoch hlbokých neurónových sietí.

Prvá konvolučná sieť s názvom Fingera, mi bola poskytnutá od firmy Innovatrics. Architektúra siete pozostáva zo štyroch konvolučných vrstiev, po ktorých nasledujú max poolingové vrstvy. Ďalej obsahuje max-out vrstvy a výstupný vektor je normalizovaný L2 normalizáciou. Architektúru siete zobrazuje obrázok 6.3.



Obr. 6.3: Architektúra konvolučnej siete Fingera.

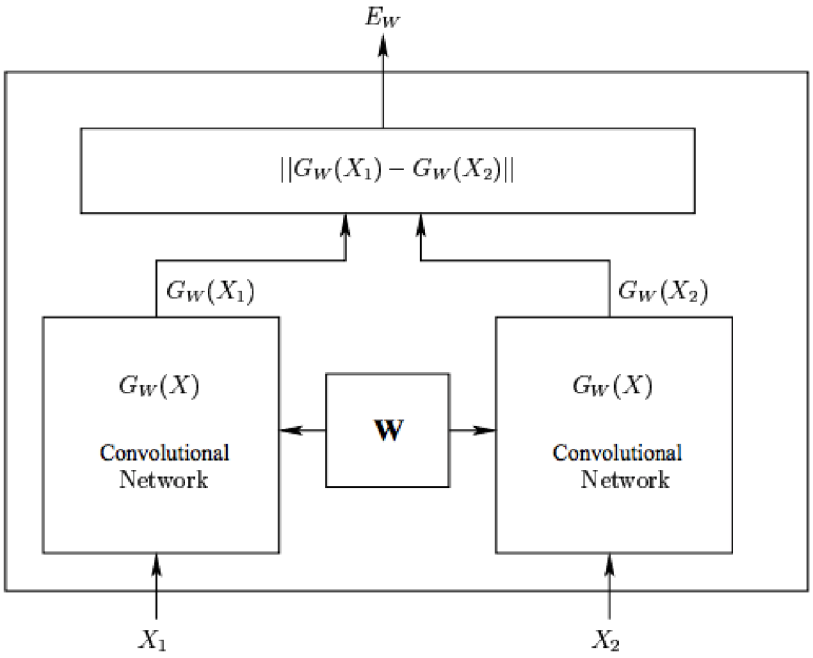
Druhú sieť som zvolil známy model VGG-Face, ktorý je založený na architektúre konvolučnej neurónovej siete VGG-Very-Deep-16, ktorá bola popísana Omkarom M. Parkhivim a kol. [42] z ústavu vizuálnej geometrie z Oxfordskej univerzity. Sieť bola natrénovaná na datasete VGG-Face, ktorý obsahuje 2.6 miliónov fotografií. VGG sieť je verejne dostupná na stránkach Model Zoo [1].

Predpokladom vhodnej siete pre extrakciu príznakových vektorov je sieť, ktorá produkuje pre rovnaké subjekty podobné výstupné vektory a menej podobné pre subjekty rozdielne. To je možné docieliť pri trénovaní siete použitím napríklad konceptu siamskej siete [12]. Počas trénovania tieto siete zdieľajú váhy 6.4. Zo vstupnej dvojice subjektov pomocou chybovej loss funkcie, napríklad Contrastive loss, vieme dosiahnuť požadovanú vlastnosť.

Ďalšia podobná technika je použitie troch sietí, tzv. triplets [46], ktoré majú na vstupe tri snímky. Dve patriace rovnakému subjektu a tretia snímka s odlišným subjektom. Využíva sa triplet-loss chybová funkcia. Požadovaná vlastnosť tejto siete je aby vektory extrahované zo snímok rovnakého subjektu mali menšiu vzdialenosť medzi sebou ako voči vektoru extrahovaného zo snímky odlišného subjektu.

Obe siete použité v tejto práci boli testované na LFW benchmarku, kde dosiahli presnosť vyše 98 percent.

[1] Model Zoo je Github repozitár, ktorý poskytuje natrénované modely neurónových sietí



Obr. 6.4: Schéma siamskej siete. Prebraté z [12].

K ohodnoteniu vstupných snímkov som použil neurónovú sieť, ktorá popisuje 68 orientačných bodov na ľudskej tvári, náklon, natočenie a uhol čeľusti.

Táto sieť je ResNet model s 27 konvolučnými vrstvami. Je založená na verzií siete ResNet-34 z práce [58], odstránením niekoľkých vrstiev a znížením počtu filtrov o polovicu. Trénovaná bola na datasete obsahujúcom vyše 3 milióny tvárí, ktorý bol zložený z viacerých datasetov ako FaceScrub [40] VGG dataset [42] a z veľkého počtu snímkov stiahnutých z internetu. Dodatočne bola dotrénovaná na datasete ibug 300-W [44]. Výstup danej siete je zobrazený graficky na obrázku 6.5.

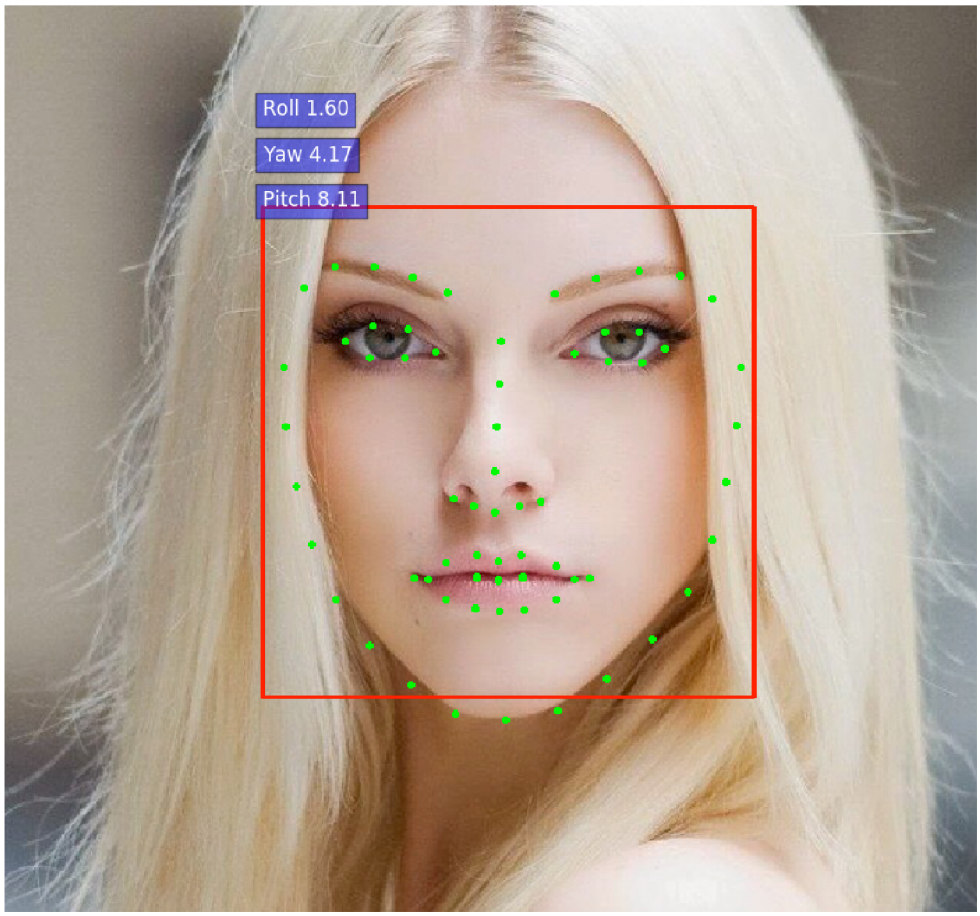
6.3 Použité nástroje

Framework V dnešnej existuje viacero frameworkov pre prácu s neurónovými sieťami. Medzi najznámejšie patria TensorFlow, Theano, Torch, MXNet či Caffe, ktorý využívam v tejto práci.

Caffe [22], čo je akronym pre Convolutional Architecture for Fast Feature Embedding, je open source framework vyvinutý na univerzite v Berkeley. Je napísaný v jazyku C++ s rozhraním v Pythone. Poskytuje možnosť trénovania pomocou grafických kariet s podporou CUDA a CuDNN pre akceleráciu výpočtov hlbokých neurónových sietí.

V tejto práci som využíval skriptovací jazyk Python a niekoľko knižníc medzi, ktoré patrí aj knižnica openCV [20] a dlib [26] pre prácu s obrazom.

Metacentrum Nakoľko práca s hlbokým učením je veľmi náročná na výpočenú kapacitu, väčšina výpočtov bola vykonávaná na serveroch Metacentra. Metacentrum je združenie CESNET primárne venované prevádzke gridovej infraštruktúry v Českej Republike. Prevádzkuje a spravuje distribuovanú výpočtovú infraštruktúru skladajúcu sa z vlastných



Obr. 6.5: Ukážka výstupu zo siete.

aj zverených výpočtových a úložných kapacít akademických centier. Poskytuje bezplatné využívanie výpočtovej a dátovej kapacity pre študentov a akademických pracovníkov.

Výpočty prebiehali na clusteroch s názvom Zubat a Doom, ktoré ponúkali vhodné prostredie pre prácu s neurónovými sieťami. Tieto clustre obsahovali grafické karty nVidia Tesla K20 s podporou CUDA a možnosťou pripojiť alebo doinštalovať potrebné knižnice ako CuDNN, openCV, numpy a framework caffe.

V rámci tejto práce bolo využitých niekoľko desiatok dní výpočtového času a pár terabajtov dátového úložiska pre niekoľko miliónov súborov.

6.4 Postup práce

Predtým než môžeme začať implementovať agregačný modul, potrebujeme mať k dispozícií vstupné dáta, teda vektory príznakov. K získaniu príznakových vektorov som postupoval podľa všeobecných postupov rozpoznávania tváre, ktoré môžeme vidieť aj na obrázku 3.1. Vstupné dáta, teda snímky, je vhodné zarovnať pre potreby určitej siete k dosiahnutiu lepších výsledkov. Táto fáza je nazývaná predspracovanie dát.

V prípade siete Fingera, sieť očakáva na vstupe čiernobiele snímky veľkosti 128x128 pixelov s fixnou pozíciou očí. Konkrétne ľavé oko na pozícií (84, 44) pixelov a pravé oko na pozícií (44, 44) pixelov. Zo vstupných snímok boli orezané tváre na základe hodnôt

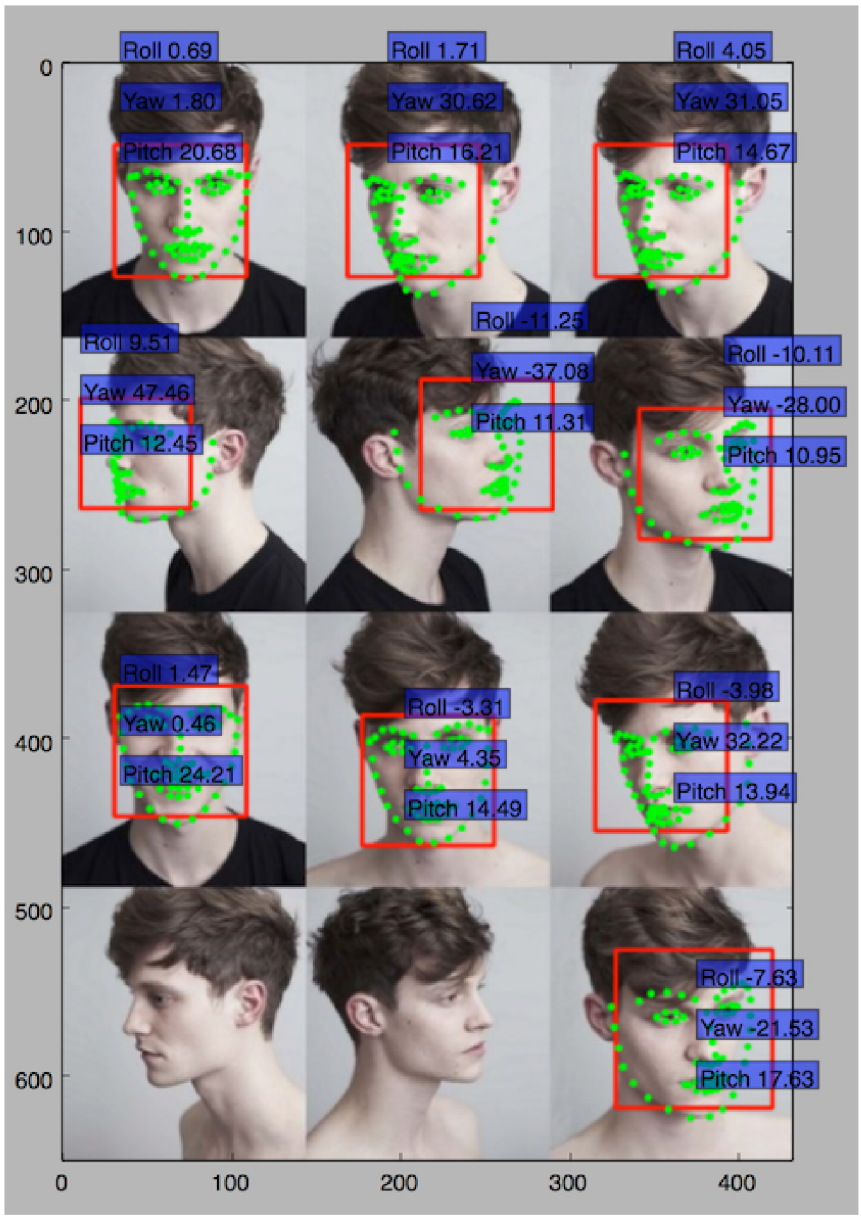
ohraničujúceho boxu z anotácií poskytnutej k datasete. Následne bola snímka prevedená do čiernobieleho formátu a jej rozmer bol zmenený na požadovanú hodnotu. Zarovnanie bolo vykonané pomocou 2D afinnej transformácie.

Takto predspracovaný snímok je pripravený pre sieť Fingera. Ešte pred extrahovaním samotných príznakov, snímok je zavedený na vstup siete, ktorá vyhodnotí pózu a natočenie tváre. Túto informáciu neskôr využije agregačný modul. Po extrahovaní príznakového vektoru zo snímky pomocou Fingera siete, spolu s vektorom si uložíme aj unikátne číslo subjektu, k identifikácií vektoru a ID číslo videa a snímku. S takto pripravenými dátami môžeme implementovať agregačný modul.

Cieľom navrhnutého agregačného modulu je pri vytváraní šablón, preferovať snímky s vyššou informačnou hodnotou. To môže byť docielené práve na základe ohodnotení snímkov. Jednotlivé snímky sú ohodnocované podľa veľkosti uhlu natočenia tváre od základnej frontálnej polohy 6.6.

Výsledný vektor, šablóna, je vypočítaná váženým priemerom. Dosiahnuté výsledky tejto agregácie sa nachádzajú v kapitole s experimentami 7. Vyhodnotenie je realizované pomocou ROC kriviek.

ROC (Receiver Operating Characteristic) krivka je nástroj pre hodnotenie a optimalizáciu binárneho klasifikačného systému (testu), ktorý ukazuje vzťah medzi špecifickosťou a senzitivitou daného testu alebo detektora pre všetky prípustné hodnoty prahu. Metrika vzdialenosti bola vypočítaná pomocou kosínusovej podobnosti.



Obr. 6.6: Ukážka ohodnocovania snímok.

Kapitola 7

Experimenty a výsledky

Táto kapitola popisuje vykonané experimenty s použitím neurónových sietí popísaných v podkapitole 6.2. Použité boli datasety IJB-A [56], UMDFaces [4] s benchmarkmi a vybraná podmnožina dát z UMDFaces/UMDFaces-videos. Agregácia je implementovaná postupom navrhnutým v 6.1. Vyhodnotenie je realizované pomocou ROC kriviek.

IJB-A verifikačný benchmark protokolu 1:1 Tento experiment skúma či navrhnutý spôsob agregácie 6.1 ovplyvní presnosť verifikačného testu. Výsledná ROC krivka 7.1 znázorňuje výsledok priemeru desiatich behov. Každý beh obsahuje okolo 11 748 testovaných párov. Modrá krivka, BA Fingera (basic average), označuje presnosť verifikácie, pri ktorej boli príznakové vektory agregované obyčajným aritmetickým priemerom. Červená krivka, WA Fingera (weighted average), vykazuje mierne zlepšenie presnosti. Príznakové vektory boli agregované navrhnutým spôsobom. Pre tento experiment boli vektory extrahované Fingera sieťou.

Pre porovnanie výsledkov bol vykonaný rovnaký experiment, len s tým rozdielom, že pre extrakciu vektorov bola použitá VGG sieť 7.2

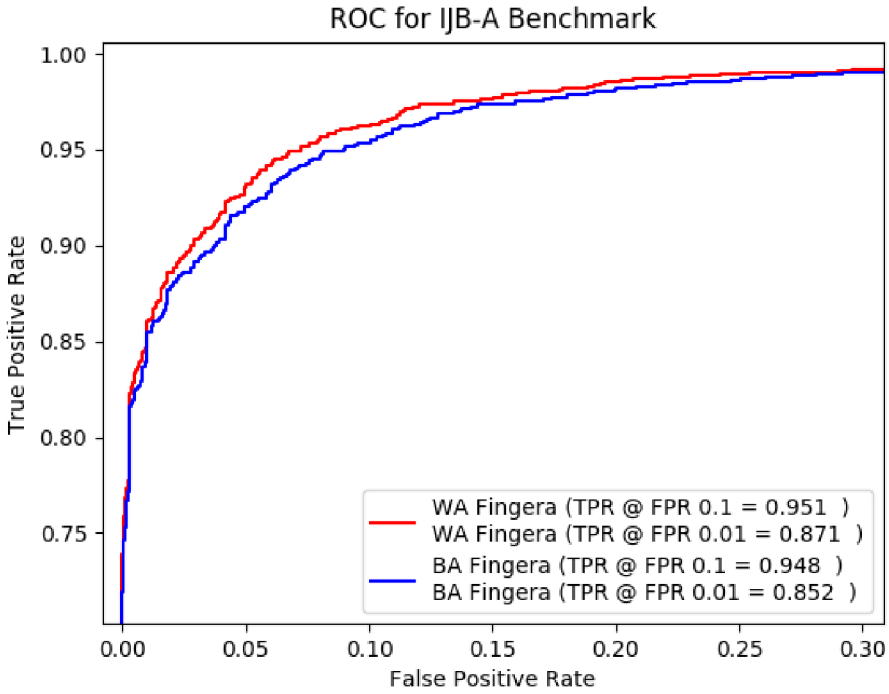
Z výsledkov oboch experimentoch môžeme vidieť, že aplikovaním navrhnutej agregácie sa zvýšila presnosť verifikácie. V dôsledku čoho môžeme usúdiť, že snímky s extrémnym natočením tváre neprospievajú ku kvalite výslednej šablóny.

Krivka 7.3 porovnáva výsledky z vykonaných experimentov s niektorými state-of-the-art metódami. Presnosť ostatných metód nájdeme v tabuľke 7.4.

UMDFaces verifikačný benchmark Dataset UMDFaces popisuje tri protokoly náročnosti evaluácie. Každý popisuje 100 000 párov subjektov na porovnávanie. Líšia sa vybranými fotografiami na ktorých sú subjekty s minimálnym natočením tváre, miernym a maximálnym natočením.

UMDFaces benchmark patrí medzi náročnejšie verifikačné protokoly čo dokazuje aj výsledok experimentu 7.5. Keď sa pozrieme na ukážky z benchmarku 4.5, tak môžeme vidieť, že aj v najľahšej variante sa nachádzajú tváre s veľkým stupňom natočenia.

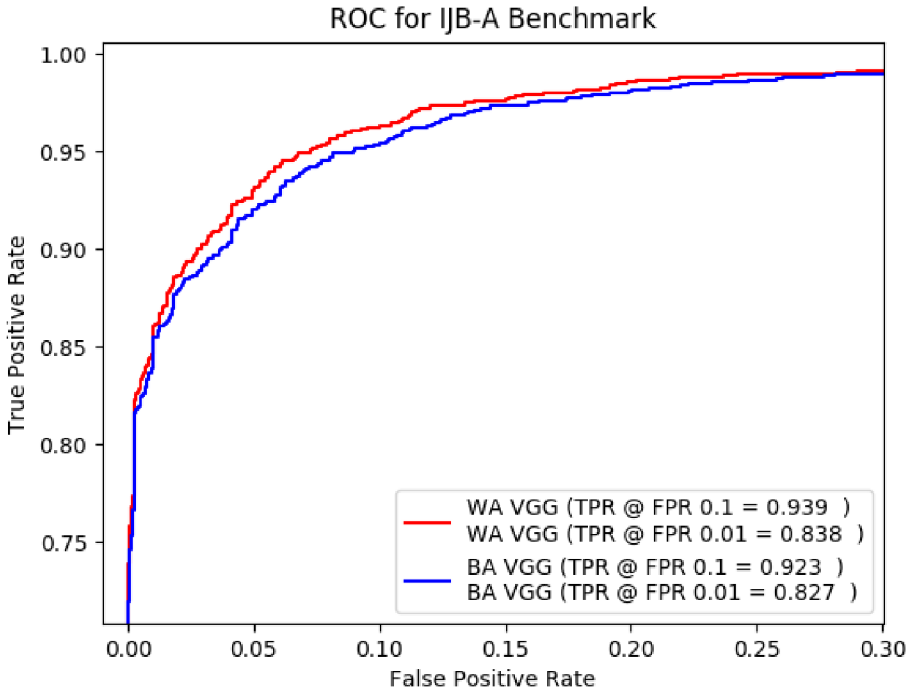
Ďalším študovaným predmetom bol typ vstupných dát k vytvorení šablón na úspešnosť verifikácie tvárí. Porovnávali sa šablóny tvorené len z fotografií, tvorené len zo snímok z jedného videa, a šablóny tvorené zmiešaním fotografií a snímkov z videa. K tomu účelu bola vytvorená metrika, podmnožina dát, z datasetov UMDFaces a UMDFaces-Videos. Veľkou výhodou je, že dataset UMDFaces-Videos obsahuje subjekty, ktoré sú podmnožinou subjektov z datasetu UMDFaces, teda subjekty sú zachytené na dostatočnom množstve fotografií a videí.



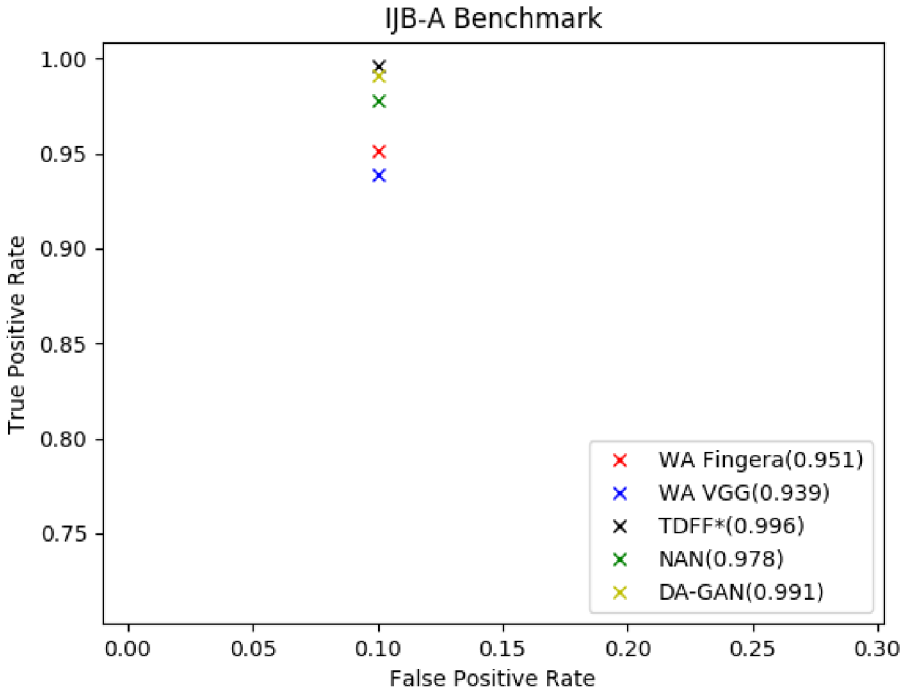
Obr. 7.1: Porovnanie metód agregácie na IJB-A benchmarku. Vektory boli extrahované Fingera sieťou.

Zo spomínaných datasetov bolo vybraných 1000 subjektov, ktoré mali aspoň 20 fotografií a 20 snímkov z jedného videa. Náhodne vygenerované dvojice 500 pozitívnych a 500 negatívnych dvojíc. Šablóny pre dané subjekty boli vypočítané z náhodného počtu fotografií alebo snímkov v rozmedzí 5-15. Porovnávala sa úspešnosť šablón vypočítaných z fotografií, snímkov a zo zmesou fotografií a snímkov.

Z krivky 7.6 najlepšie vyšli šablóny, ktoré boli tvorené fotografiami a snímkami. Najslabší výsledok dosiahli šablóny tvorené len snímkami z videa.



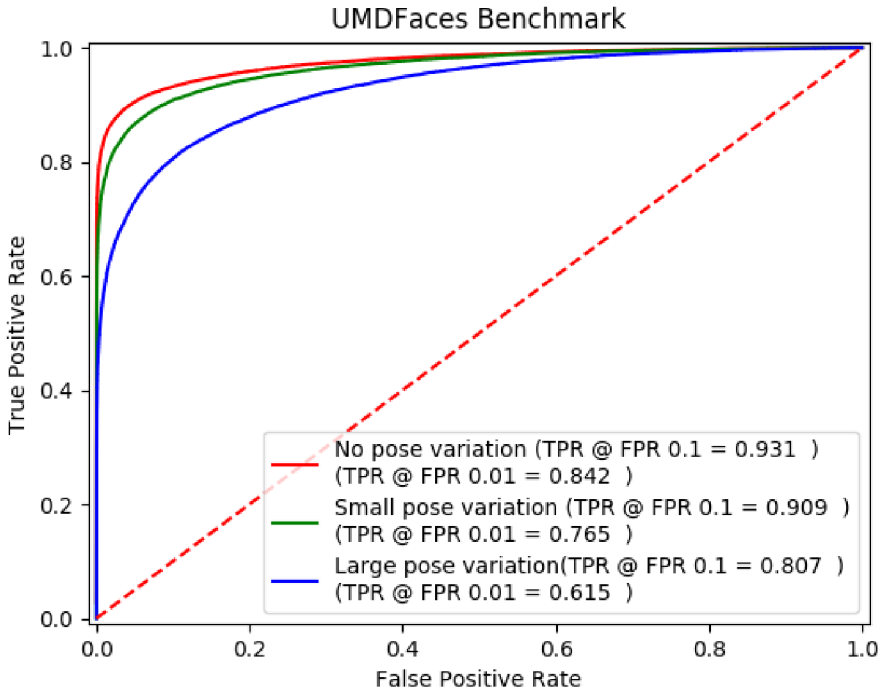
Obr. 7.2: Úspešnosť siete VGG na IJB-A benchmarku.



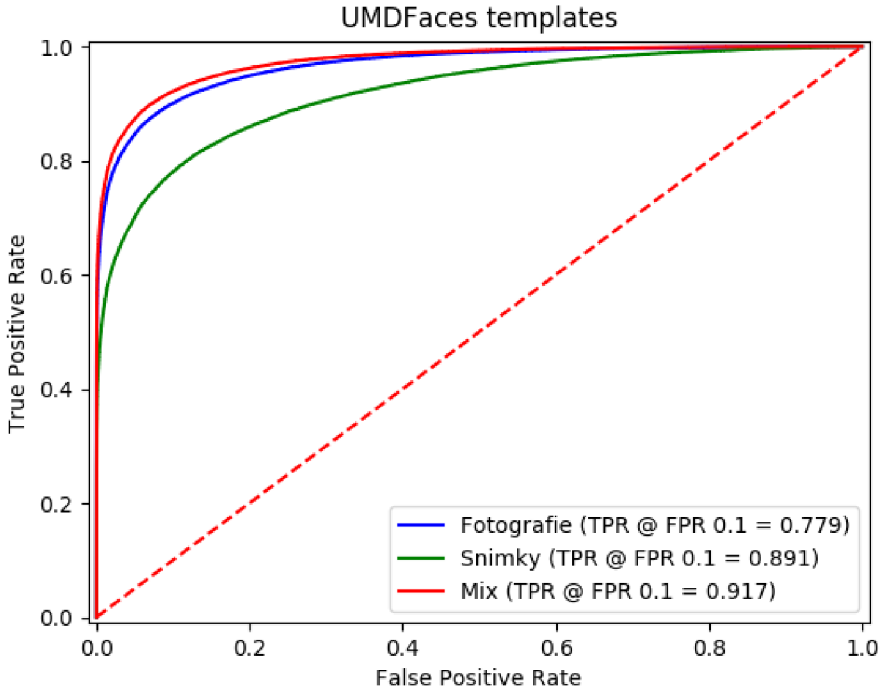
Obr. 7.3: Porovnanie dosiahnutej presnosti s metódami state-of-the-art.

Method	1:1 Verification TAR		
	FAR=0.001	FAR=0.01	FAR=0.1
OpenBR	0.104±0.014	0.236±0.009	0.433±0.006
GOTS	0.198±0.008	0.406±0.014	0.627±0.012
B-CNN	-	-	-
Pooling faces	-	0.309	0.631
LSFS	0.514±0.060	0.733±0.034	0.895±0.013
Deep Multi-pose	-	0.787	0.911
DCNN*manual*+metric	-	0.787±0.043	0.947±0.011
Triplet Similarity	0.590±0.050	0.790±0.030	0.945±0.002
VGG-Face	-	0.805±0.030	-
PAMs	0.652±0.037	0.826±0.018	-
DCNN*fusion*	-	0.838±0.042	0.967±0.009
FF-GAN	0.663±0.033	0.852±0.010	-
DR-GAN*fuse*	0.699±0.029	0.831±0.017	-
Masi *et al.*	0.725	0.886	-
Triplet Embedding	0.813±0.020	0.900±0.010	0.964±0.005
Template Adaptation	0.836±0.027	0.939±0.013	0.979±0.004
Chen *et al.*	0.760±0.038	0.889±0.016	0.968±0.005
All-In-One+TPE	0.823±0.020	0.922±0.010	0.976±0.004
NAN	0.881±0.011	0.941±0.008	0.978±0.003
Hayat *et al.*	-	-	-
DA-GAN	0.930±0.005	0.976±0.007	0.991±0.003
L_2-softmax	0.938±0.008	0.968±0.004	0.987±0.002
L_2-softmax +TPE	0.943±0.005	0.970±0.004	0.984±0.002
TDFF	0.919±0.006	0.961±0.007	0.988±0.003
TDFF+TPE	0.921±0.005	0.961±0.007	0.989±0.003
TDFF*	**0.979±0.004**	**0.991±0.002**	**0.996±0.001**

Obr. 7.4: Dosiahnutá presnosť ostatných state-of-the-art metód. Prebraté z [59]



Obr. 7.5: Dosiahnutá úspešnosť na UMDFaces verifikačnom benchmarku.



Obr. 7.6: ROC krivka obsahujúca všetky tri experimenty.

Kapitola 8

Záver

Prvá polovica práce stručne prechádza historickým vývojom neurónových sietí k popisu všeobecného modelu neurónu a architektúry siete. Vysvetlený je proces učenia, základné funkcie konvolučných sietí a postup rozpoznávanie tvárí. Predstavené sú tiež dostupné datasety a rozdiely medzi nimi.

V rámci tejto práce bol navrhnutý spôsob agregácie príznakových vektorov zo snímkov z videa. Motiváciou skúmania metód agregácie je rozpoznávanie tvárí z video záznamu, ktorý obsahuje väčšie množstvo informácií než statická fotografia. Navrhnutý spôsob skúma kvalitu vytvorených šablón vzhľadom k snímkam, z ktorých je vypočítaný. Ohodnocuje jednotlivé snímky váhami, ktorými budú ich príznakové vektory násobené. Tým agregačný modul dokáže kontrolovať ako veľmi daný snímok ovplyvní výslednú šablónu.

Experimentami sa podarilo dokázať zvýšenie presnosti na dvoch verifikačných benchmarkoch. Ďalej v práci bolo zistené, že šablóny tvorené fotografiami aj snímkami, dosahujú najlepších výsledkov.

V rámci budúcej práce by bolo zaujímavé skúmať vplyv rôznych veľkostí výstupných šablón, 3D zarovnania tváre alebo použitie rekurentných sietí.

Literatúra

[1] Ahonen, T.; Hadid, A.; Pietikäinen, M.; aj.: Face recognition with local binary patterns. *Proc. of the European Conference on Computer Vision (ECCV)*, 2004: s. 469–481, ISSN 03029743, doi:10.1007/978-3-642-25449-9_2.
URL http://www.springerlink.com/index/P5D9XP9GFKEX5GK9.pdf

[2] Attneave, F.; B., M.; Hebb, D. O.: The Organization of Behavior; A Neuropsychological Theory. *The American Journal of Psychology*, ročník 63, č. 4, 1950: str. 633, ISSN 00029556, doi:10.2307/1418888.
URL http://www.jstor.org/stable/1418888?origin=crossref

[3] Bansal, A.; Castillo, C.; Ranjan, R.; aj.: The Do's and Don'ts for CNN-based Face Verification. 2017, 1705.07426.
URL http://arxiv.org/abs/1705.07426

[4] Bansal, A.; Nanduri, A.; Castillo, C.; aj.: UMDFaces: An Annotated Face Dataset for Training Deep Networks. 2016, 1611.01484.
URL http://arxiv.org/abs/1611.01484

[5] Beveridge, J. R.; Givens, G. H.; Scruggs, W. T.; aj.: The Challenge of Face Recognition from Digital Point-and-Shoot Cameras. 2013.
URL
http://biometrics.nist.gov/cs{_}links/face/PaSC/pasc2013{_}NISTIR.pdf

[6] Briefs, S.; Applied, I. N.: *SPRINGER BRIEFS IN APPLIED SCIENCES AND An Introduction to Neural Network Methods for Differential Equations*. 2015, ISBN 9789401798150.
URL https://universalflowuniversity.com/Books/ComputerProgramming/NeuralNetworksandDeepLearning/AnIntroductiontoNeuralNetworkMethodsforDifferentialEquations.pdf

[7] Cao, Q.; Shen, L.; Xie, W.; aj.: VGGFace2: A dataset for recognising faces across pose and age. 2017, 1710.08092.
URL http://arxiv.org/abs/1710.08092

[8] Chen, B. C.; Chen, C. S.; Hsu, W. H.: Cross-age reference coding for age-invariant face recognition and retrieval. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, ročník 8694 LNCS, č. PART 6, 2014: s. 768–783, ISSN 16113349, doi:10.1007/978-3-319-10599-4_49.
URL http://cmlab.csie.ntu.edu.tw/{~}sirius42/papers/chen14eccv.pdf

[9] Chen, D.; Cao, X.; Wang, L.; aj.: Bayesian Face Revisited: A Joint Formulation. , č. 1, 2012: s. 566–579, doi:10.1007/978-3-642-33712-3_41.
URL http://link.springer.com/10.1007/978-3-642-33712-3{_}41

[10] Chen, D.; Ren, S.; Wei, Y.; aj.: Joint Cascade Face Detection and Alignment. 2014: s. 109–122, doi:10.1007/978-3-319-10599-4_8.
URL http://link.springer.com/10.1007/978-3-319-10599-4{_}8

[11] Chen, J. C.; Patel, V. M.; Chellappa, R.: Unconstrained face verification using deep CNN features. In *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016, ISBN 9781509006410, doi:10.1109/WACV.2016.7477557, 1508.01722.

[12] Chopra, S.; Hadsell, R.; Y., L.: Learning a similiarty metric discriminatively, with application to face verification. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005: s. 349–356, ISSN 10636919, doi:10.1109/CVPR.2005.202.
URL http://yann.lecun.com/exdb/publis/pdf/chopra-05.pdf

[13] Dettmers, T.: Deep Learning in a Nutshell: Core Concepts. https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-core-concepts, 2015, [Online; navštívené cit.2017-12-05].

[14] Diego, S. A. N.: 862 18 120,. , č. V, 1985.
URL http://www.dtic.mil/dtic/tr/fulltext/u2/a164453.pdf

[15] Eringen, C. A.: Unclassified Limitation Changes To : From :. *Distribution*, 1973: s. 1–35.
URL http://www.dtic.mil/dtic/tr/fulltext/u2/241531.pdf

[16] Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, ročník 36, č. 4, 1980: s. 193–202, ISSN 03401200, doi:10.1007/BF00344251, arXiv:1011.1669v3.
URL https://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf

[17] Guo, Y.; Zhang, L.; Hu, Y.; aj.: MS-Celeb-1M: A Dataset and Benchmark for Large Scale Face Recognition. *CoRR*, ročník abs/1607.0, 2016: s. 1–17, ISSN 2470-1173, doi:1607.08221, 1607.08221.
URL https://www.microsoft.com/en-us/research/wp-content/uploads/2016/08/MSCeleb-1M-a.pdf

[18] Gupta, S.: A short introduction to heavy-ion physics. *Journal of Japanese Society for Artificial Intelligence*, ročník 14, č. 5, 2015: s. 771–780, ISSN 10450823, doi:citeulike-article-id:765005, 1508.01136.
URL http://arxiv.org/abs/1508.01136

[19] Huang, G. B.; Ramesh, M.; Berg, T.; aj.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *University of Massachusetts Amherst Technical Report*, ročník 1, 2007: s. 07–49, ISSN 1996756X, doi:10.1.1.122.8268.

[20] Itseez: Open Source Computer Vision Library. https://github.com/itseez/opencv, 2015.

[21] Jia, Y.: *Fine-tuning CaffeNet for Style Recognition on "Flickr Style" Data.* http://caffe.berkeleyvision.org/gathered/examples/finetune_flickr_style.html, 2016, [Online; navštívené 17.12.2017].

[22] Jia, Y.; Shelhamer, E.; Donahue, J.; aj.: Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[23] Karpathy, A.: CS231n Convolutional Neural Networks for Visual Recognition. http://cs231n.github.io/neural-networks-1/, 2015, [Online; navštívené 11.10.2017].

[24] Karpathy, A.: CS231n Convolutional Neural Networks for Visual Recognition. https://cs231n.github.io/convolutional-networks/, 2015, [Online; navštívené17.12.2016].

[25] Kemelmacher-Shlizerman, I.; Seitz, S.; Miller, D.; aj.: The MegaFace Benchmark: 1 Million Faces for Recognition at Scale. 2015, ISSN 10636919, doi:10.1109/CVPR.2016.527, 1512.00596.
URL http://arxiv.org/abs/1512.00596

[26] King, D. E.: Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, ročník 10, 2009: s. 1755–1758, ISSN 15324435, doi:10.1145/1577069.1755843.
URL http://jmlr.csail.mit.edu/papers/v10/king09a.html

[27] Klare, B. F.; Klein, B.; Taborsky, E.; aj.: Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ročník 07-12-June, 2015, ISBN 9781467369640, ISSN 10636919, s. 1931–1939, doi:10.1109/CVPR.2015.7298803.

[28] Krizhevsky, A.; Sutskever, I.; Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 2012: s. 1–9, ISSN 10495258, doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007, 1102.0183.

[29] Kumar, N.; Berg, A. C.; Belhumeur, P. N.; aj.: Attribute and simile classifiers for face verification. *Proceedings of the IEEE International Conference on Computer Vision*, 2009: s. 365–372, ISSN 1550-5499, doi:10.1109/ICCV.2009.5459250.
URL http://www.cs.columbia.edu/CAVE/publications/pdfs/Kumar{_}ICCV09.pdf

[30] LeCun, Y.: A theoretical framework for Back-Propagation. 1988, doi:10.1007/978-3-642-35289-8, arXiv:1011.1669v3.
URL http://yann.lecun.com/exdb/publis/pdf/lecun-88.pdf

[31] LeCun, Y.; Bottou, L.; Bengio, Y.; aj.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, ročník 86, č. 11, 1998: s. 2278–2323, ISSN 00189219, doi:10.1109/5.726791, 1102.0183.

[32] Li, Haoxiang and Lin, Zhe and Shen, Xiaohui and Brandt, Jonathan and Hua, G.: A Convolutional Neural Network Approach for Face Detection. *Cvpr*, 2015: s. 5325–5334, ISSN 1063-6919, doi:10.1109/CVPR.2015.7299170.
URL http://users.eecs.northwestern.edu/{~}xsh835/assets/cvpr2015{_}cascnn.pdf

[33] Liao, S.; Jain, A. K.; Li, S. Z.: A Fast and Accurate Unconstrained Face Detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ročník 38, č. 2, 2016: s. 211–223, ISSN 01628828, doi:10.1109/TPAMI.2015.2448075, 1408.1656.
URL https://arxiv.org/pdf/1408.1656.pdf

[34] Lipton, Z. C.; Berkowitz, J.; Elkan, C.: A Critical Review of Recurrent Neural Networks for Sequence Learning. 2015: s. 1–38, ISSN 9781450330633, doi:10.1145/2647868.2654889, 1506.00019.
URL http://arxiv.org/abs/1506.00019

[35] Masi, I.; Rawls, S.; Medioni, G.; aj.: Pose-Aware Face Recognition in the Wild. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, s. 4838–4846, doi:10.1109/CVPR.2016.523.

[36] Mathias, M.; Benenson, R.; Pedersoli, M.; aj.: Face detection without bells and whishes. *European Conference on Computer Vision (ECCV)*, 2014: s. 720–735, ISSN 16113349, doi:10.1007/978-3-319-10593-2_47.
URL https://scholar.google.co.nz/citations?view{_}op=view{_}citation{&}continue=/scholar?hl=en{&}start=20{&}as{_}sdt=0,5{&}scilib=1{&}citilm=1{&}citation{_}for{_}view=14heRRwAAAAJ:UebtZRa9Y70C{&}hl=en{&}oi=p

[37] McCarthy, J.; Minsky, M. L.; Rochester, N.; aj.: A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, ročník 27, č. 4, 1955: s. 12–14, ISSN 0738-4602, doi:http://dx.doi.org/10.1609/aimag.v27i4.1904, 9809069v1.
URL http://www.aaai.org/ojs/index.php/aimagazine/article/view/1904{%}5Cnhttp://www.mendeley.com/catalog/proposal-dartmouth-summer-research-project-artificial-intelligence-august-31-1955/{%}5Cnhttp://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

[38] McCulloch, W. S.; Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, ročník 5, č. 4, 1943: s. 115–133, ISSN 00074985, doi:10.1007/BF02478259, arXiv:1011.1669v3.
URL https://pdfs.semanticscholar.org/5272/8a99829792c3272043842455f3a110e841b1.pdf

[39] Minsky, M.; Papert, S.: *Perceptrons - an introduction to computational geometry*. MIT Press, 1987.

[40] Ng, H. W.; Winkler, S.: A data-driven approach to cleaning large face datasets. *2014 IEEE International Conference on Image Processing, ICIP 2014*, 2014: s. 343–347, ISSN 1522-4880, doi:10.1109/ICIP.2014.7025068.
URL http://vintage.winklerbros.net/Publications/icip2014a.pdf

[41] Nielsen, M. A.: *Neural Networks and Deep Learning*. Determination Press, 2015.
URL http://neuralnetworksanddeeplearning.com/index.html

[42] Parkhi, O. M.; Vedaldi, A.; Zisserman, A.: Deep Face Recognition. In *Proceedings of the British Machine Vision Conference (BMVC)*, editácia M. W. J. Xianghua Xie; G. K. L. Tam, BMVA Press, September 2015, ISBN 1-901725-53-7, s. 41.1–41.12, doi:10.5244/C.29.41.
URL https://dx.doi.org/10.5244/C.29.41

[43] Rowley, H.; Baluja, S.; Kanade, T.: Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ročník 20, č. 1, 1998: s. 23–38, ISSN 0162-8828, doi:10.1109/34.655647.
URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.2367{&}rep=rep1{&}type=pdf

[44] Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; aj.: 300 Faces In-The-Wild Challenge: database and results. *Image and Vision Computing*, ročník 47, 2015: s. 3–18, ISSN 02628856, doi:10.1016/j.imavis.2016.01.002.
URL http://dx.doi.org/10.1016/j.imavis.2016.01.002https://ibug.doc.ic.ac.uk/media/uploads/documents/sagonas{_}2016{_}imavis.pdf

[45] Sangeetha, Y.; Latha, P. M.; Narasimham, C.; aj.: Face Detection using SMQT Techniques. *International Journal of Computer Science and Engineering Technology*, ročník 2, č. 1, 2012: s. 1780–1783.
URL http://ijcset.net/docs/Volumes/volume2issue1/ijcset2012020104.pdf

[46] Schroff, F.; Kalenichenko, D.; Philbin, J.: FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ročník 07-12-June, 2015, ISBN 9781467369640, ISSN 10636919, s. 815–823, doi:10.1109/CVPR.2015.7298682, 1503.03832.

[47] Sharma, P.; Yadav, R. N.; Arya, K. V.: Face recognition from video using generalized mean deep learning neural network. In *2016 4th International Symposium on Computational and Business Intelligence (ISCBI)*, Sept 2016, s. 195–199, doi:10.1109/ISCBI.2016.7743283.

[48] Simonyan, K.; Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014: s. 1–14, ISSN 09505849, doi:10.1016/j.infsof.2008.09.005, 1409.1556.
URL http://arxiv.org/abs/1409.1556

[49] Sun, Y.; Liang, D.; Wang, X.; aj.: DeepID3: Face Recognition with Very Deep Neural Networks. 2015: s. 2–6, 1502.00873.
URL http://arxiv.org/abs/1502.00873

[50] Sun, Y.; Wang, X.; Tang, X.: Deeply learned face representations are sparse, selective, and robust. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ročník 07-12-June-2015, 2015: s. 2892–2900, ISSN 10636919, doi:10.1109/CVPR.2015.7298907, 1412.1265.
URL https://arxiv.org/pdf/1412.1265.pdf

[51] Szegedy, C.; Liu, W.; Jia, Y.; aj.: Going Deeper with Convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.
URL http://arxiv.org/abs/1409.4842

[52] Taigman, Y.; Yang, M.; Ranzato, M.; aj.: DeepFace: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, ISBN 9781479951178, ISSN 10636919, s. 1701–1708, doi:10.1109/CVPR.2014.220, 1501.05703.

[53] Taigman, Y.; Yang, M.; Ranzato, M. A.; aj.: Web-Scale Training for Face Identification | Publications | Research at Facebook | Facebook. , č. i, 2015: str. 2, 1406.5266.
URL https://research.facebook.com/publications/787935884654205/web-scale-training-for-face-identification/?utm{_}source=researchdot{_}rss{_}feed{&}utm{_}medium=rss{&}utm{_}campaign=RSS+Feed

[54] TuringFinance: 10 misconceptions about Neural Networks. 2014, [Online; navštívené 13-September-2017].
URL http://www.turingfinance.com/misconceptions-about-neural-networks/

[55] Viola, P.; Jones, M. J.: Robust Real-time Object Detection. *International Journal of Computer Vision*, , č. February, 2001: s. 1–30, ISSN 09205691, doi:10.1.1.23.2751.
URL http://www.hpl.hp.com/techreports/Compaq-DEC/CRL-2001-1.pdf

[56] Whitelam, C.; Taborsky, E.; Blanton, A.; aj.: IARPA Janus Benchmark-B Face Dataset. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, ročník 2017-July, 2017: s. 592–600, ISSN 21607516, doi:10.1109/CVPRW.2017.87.
URL https://pdfs.semanticscholar.org/4c1a/decae35d53e3a377b8de0a49b3cdd6960907.pdf

[57] Wolf, L.; Hassner, T.; Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, ISBN 9781457703942, ISSN 10636919, s. 529–534, doi:10.1109/CVPR.2011.5995566.

[58] Wu, S.; Zhong, S.; Liu, Y.: Deep residual learning for image steganalysis. *Multimedia Tools and Applications*, 2017: s. 1–17, ISSN 15737721, doi:10.1007/s11042-017-4440-4, 1512.03385.
URL https://arxiv.org/pdf/1512.03385.pdf

[59] Xiong, L.; Karlekar, J.; Zhao, J.; aj.: A Good Practice Towards Top Performance of Face Recognition: Transferred Deep Feature Fusion. *ArXiv e-prints*, Apríl 2017, 1704.00438.

[60] Yang, J.; Ren, P.; Chen, D.; aj.: Neural Aggregation Network for Video Face Recognition. mar 2016, 1603.05474.
URL http://arxiv.org/abs/1603.05474

[61] Yang, S.; Luo, P.; Loy, C. C.; aj.: WIDER FACE: A face detection benchmark. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, ročník 2016-Decem, 2016: s. 5525–5533, ISSN 10636919,

doi:10.1109/CVPR.2016.596, 1511.06523.
URL https:
//pdfs.semanticscholar.org/52d7/eb0fbc3522434c13cc247549f74bb9609c5d.pdf

[62] Yang, S.; Luo, P.; Loy, C.-C. C.; aj.: From Facial Parts Responses to Face Detection: A Deep Learning Approach. *2015 IEEE International Conference on Computer Vision (ICCV)*, , č. 3, 2015: s. 3676–3684, ISSN 15505499, doi:10.1109/ICCV.2015.419, 1509.06451.
URL http://arxiv.org/abs/1509.06451{%}5Cnhttp:
//ieeexplore.ieee.org/document/7410776/

Prílohy

Príloha A

Obsah priloženého pamäťového média

texty obsahuje vygenerované pdf diplomovej práce a zdrojové dokumenty

src obsahuje skripty použité v práci

README obsahuje krátky popis k súborom

video obsahuje video prezentujúcu prácu