

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



VYHLADAVANIE KONKRÉTNEJ OSOBY VO VIDEU
PRI ČIASTOČNEJ OKLÚZII TVÁRE

Diplomová práca

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



VYHLÁDÁVANIE KONKRÉTNEJ OSOBY VO VIDEU
PRI ČIASTOČNEJ OKLÚZII TVÁRE

Diplomová práca

Študijný program: Aplikovaná informatika
Študijný odbor: 2511 Aplikovaná informatika
Školiace pracovisko: Katedra aplikovanej informatiky
Školiteľ: RNDr. Zuzana Černeková, PhD.

Bratislava, 2020

Bc. Marianna Ráchelová



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Marianna Ráchelová
Študijný program: aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: aplikovaná informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Vyhľadávanie konkrétnej osoby vo videu pri čiastočnej oklúzii tváre
Search of certain specific person with partially occluded face in video sequences

Anotácia: Vyhľadanie osoby vo videozázname podľa zadaného vzoru tj. tváre človeka v prípade, že tvár je čiastočne zakrytá. Naštudovať problematiku detekcie a sledovania ľudských tvárí. Analyzovať existujúce riešenia publikované v dostupnej odbornej literatúre. Vytvoriť databázu videozáznamov a hľadaných tvárí pre testovacie účely. Navrhnuť a implementovať metódu, ktorá vyhľadá osobu podľa zadaného vizuálneho vzoru t.j. tváre človeka vo videozázname, v ktorom nie je viditeľná celá tvár. Vyhodnotiť dosiahnuté výsledky.

Cieľ: Vyhľadanie osoby vo videozázname podľa zadaného vzoru tj. tváre človeka v prípade, že tvár je čiastočne zakrytá. Naštudovať problematiku detekcie a sledovania ľudských tvárí. Analyzovať existujúce riešenia publikované v dostupnej odbornej literatúre. Vytvoriť databázu videozáznamov a hľadaných tvárí pre testovacie účely. Navrhnuť a implementovať metódu, ktorá vyhľadá osobu podľa zadaného vizuálneho vzoru t.j. tváre človeka vo videozázname, v ktorom nie je viditeľná celá tvár. Vyhodnotiť dosiahnuté výsledky.

Vedúci: RNDr. Zuzana Černeková, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.
Dátum zadania: 24.10.2016

Dátum schválenia: 03.10.2018
prof. RNDr. Roman Ďurikovič, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

Čestne prehlasujem, že diplomovú prácu s názvom: Vyhľadávanie konkrétnej osoby vo videu pri čiastočnej oklúzii tváre som vypracovala samostatne, na základe konzultácií, teoretických poznatkov, praktických poznatkov a štúdia odbornej literatúry. Neporušila som autorský zákon a zoznam použitej literatúry som uviedla na príslušnom mieste.

Bratislava, 2020

.....
Bc. Marianna Ráchelová

Pod'akovanie

Chcem sa pod'akovať svojej školiteľke RNDr.Zuzane Černekovej,PhD.,za cennú pomoc, rady, konzultácie a čas, ktorý mi venovala počas písania diplomovej práce.

Abstrakt

Slovenský abstrakt. Stručná sumarizácia výsledkov práce pochopiteľná pre bežného informatika s rozsahom jedného odstavca.

Kľúčové slová: jedno, druhé, tretie, ..

Abstract

Abstrakt preložený do angličtiny.

Keywords: ...

Obsah

Úvod	3
1 Analýza súčasných riešení	4
1.1 Súvisiace práce	5
1.2 Databázy	8
1.2.1 Labeled faces in the wild	9
1.2.2 Aleix and Robert	10
1.2.3 Face occlusion dataset	13
1.3 Konvolučné neurónové siete	15
1.3.1 Konvolučná vrstva	16
1.3.2 Pooling vrstva	18
1.3.3 Plne prepojená vrstva	19
1.4 Siamská neurónová sieť	19
1.4.1 Triplet loss	21
1.5 Facenet	21
2 Návrh riešenia	22
3 Implementácia	23
4 Výsledky	24

OBSAH

ix

Záver

25

Zoznam obrázkov

1.1	Labeled faces in the Wild[6]	10
1.2	Aleix and Robert databáza[8]	12
1.3	Rôzne oklúzie tváre[15]	13
1.4	Rôzne oklúzie tváre[15]	14
1.5	Rôzne oklúzie tváre[15]	15
1.6	Architektúra konvolučnej neurónovej siete	16
1.7	Konvolučná vrstva[3]	17
1.8	Pooling vrstva[3]	19
1.9	Siamska neurónová sieť	20

Zoznam tabuliek

Úvod

Motivácia

Detegovanie konkrétnych osôb patrí v dnešnej dobe k významným spoločenským témam. V modernej spoločnosti sa pohybujú významné osoby ako i obyčajní ľudia. Každý z nás je istým spôsobom sledovaný vo vonkajšom prostredí. V obchodných domoch sú nainštalované kamery, ktoré sledujú naše kroky a tiež bezpečnosť ľudí a pracovníkov nachádzajúcich sa v budove. Kamery inštalujeme aj na naše domy, nachádzajú sa aj vo verejnom priestore ako napríklad ulice mesta sú tiež neustále sledované kvoli premávke alebo bezpečnosti občanov. Kamerové záznamy slúžia pri vyšetrovaní kriminálnych činov, kde je dôležité detegovať zo záznamu nielen priebeh inkriminovanej trestnej činnosti, ale v najlepšom prípade detegovať konkrétnu osobu. V niektorých prípadoch patrí detekcia konkrétnej osoby vo videu k nežiadúcim udalostiam. Osoby známe v spoločnosti si neželajú byť sledované kamerami novinármi, ktorí týmto spôsobom zasahujú do ich súkromia bez ich vedomia, či povolenia.

Kapitola 1

Analýza súčasných riešení

Detekcia osôb a následne rozpoznanie konkrétnej osoby dnes patrí k významným témam počítačového videnia. Riešenie takejto úlohy je netriviálne vzhľadom k tomu, že dataset nasnímaných osôb je veľmi vnútrotriedne variabilný. Osoby na obrázkoch sú často zakryté rôznymi predmetmi či už oblečením, okuliarmi alebo ďalšou osobou prípadne iným objektom. Ak osoba na obrázku je čiastočne alebo úplne zakrytá hovoríme o oklúzií. Obrázky v datasete tiež podliehajú rôznym svetelným podmienkam alebo môžu byť zašumené. Osoby majú rôznorodý vzhľad, vek, pohlavie atď. Riešenie úlohy rozpoznania konkrétnej osoby je dôležitá v otázke bezpečnostných systémov, bezpečnosti alebo riešenie úlohy môže byť nápomocné pri vyšetrovaní kriminálnych činov. V kapitole analýza súčasných riešení si priblížime súvisiace práce na tému rozpoznávanie osôb vo videu, detekcia oklúzie tváre. Predstavíme najčastejšie používané databázy v oblasti detekcie a rozpoznávania tvárí. Priblížime si konvolučné neurónové siete a jednotlivé modely konvolučných neurónových sietí, ktoré sa aktuálne používajú v oblasti počítačového videnia. Na záver kapitoly vysvetlíme dostupné a často používané technológie pri riešení úlohy rozpoznávania konkrétnych osôb.

1.1 Súvisiace práce

V podkapitole Súvisiace práce stručne predstavíme práce , ktoré sa doteraz zaoberali detekciou osôb, detekciou tváre, detekciou oklúzie tváre a rozpoznávaním osôb. Popíšeme ako autori riešili dané problémy v oblasti počítačového videnia. Doteraz bolo predstavené veľké množstvo riešení daných problémov a tiež metód, ktoré boli pomerne presné, avšak jednotlivé riešenia sa líšia v stabilite a vypočtovej zložitosti. Čitateľovi stručne predstavíme zaujímavé riešenia daných úloh, obsiahnejšie sú popísané v citovanej literatúre.

Detekciu oklúzie tváre predstavili autori v článku Face Occlusion Detection Using Deep Convolutional Neural Networks[15]. Riešenie detekcie oklúzie tváre bolo realizované pomocou modelu konvolučných neurónových sietí. Autori použili dva modely neurónových sietí. Prvý model detegoval tvárovú oblasť z obrázku hornej polovice tela osoby a výstup, druhý model klasifikoval oklúziu v tvárovej oblasti. To znamená , že výstup z prvého modelu neurónovej siete bol použitý ako vstup do druhého modelu neurónovej siete. Výstupom z druhého modelu neurónovej siete bola zaklasifikovaná oklúzia tváre. Autori pri tréňovaní a testovaní modelu použili tri rôzne datasety a to Labeled faces in the wild[6], Aleix and Robert[8] a ich vlastný dataset Face occlusion dataset[15], ktorý vytvorili pomocou 220 osôb. Osoby boli odfotené s natočením tváre pod určitým uhlom, bez oklúzie tváre a s rôznymi oklúziami tváre. Autori dosiahli so svojím riešením úspešnosti 98.58% ,94.55% a 95.41% na datasetoch Aleix a Robert, Face occlusion datasete a Labeled faces in the wild datasete.

Riešenie opätovného identifikovania osoby vo videu predstavili autori v článku Video person re-identification using learned clip similarity[9]. Video sekvenciu zachytávajúcu konkrétnu osobu vo videu rozdelili na viacero častí tzv. klipy a následne agregovali tieto klipy podľa podobností. Vo video kli-

poch zohľadňovali rôzne pózy osoby, oklúziu osoby alebo rozmazanie pohybu osoby. Pre výpočet podobností v pároch klipov použili scoring funkciu. Nízke hodnoty scoring funkcie dosiahli páry klipov, kde osoba je zachytená v odlišných pózach. Naopak, páry klipov, kde osoby vykazovali podobné pózy, dosiahli vyššie hodnoty scoring funkcie. Metóda sa naučí váhy dôležitosti jednotlivých klipov. Autori použili na riešenie tohto problému model neurónovej siete 3D CNN. Videozáznamy sa najprv rozdelia na klipy, ktoré sa skombinujú, aby sa získali príznaky klipu. Kombinované príznaky klipu sa potom spoja s dôležitým skóre ako váha. Finálny združený reprezentačný vektor sa potom použije na výpočet podobnosti. Autori použili svoju metódu na viacerých video datasetoch menovite MARS, DukeMTMC-VideoReID a PRID2011, ktoré sú bližšie popísané a odkazované v článku.

Metóda na riešenie rozpoznávania tváre s oklúziou bola predstavená aj v článku Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network[11]. Autori hľadali riešenie pre problém rozpoznávania tváre za rôznych podmienok ako napríklad rôzne pózy, výrazy tváre, osvetlenie a oklúzia. Autori boli inšpirovaní ľudským vizuálnym vnímaním tváre, t.j. ľudia často filtrujú také časti tváre, ktoré sú zakryté. Preto navrhli odstraňovať príznakové prvky, ktoré obašujú oklúziu. Autormi navrhovaný framework je založený na základnom modeli konvolučnej neurónovej siete trénovanom na rozpoznávanie tváre. Riešením úlohy rozpoznávania zakrytej tváre je model Siamskej neurónovej siete, ktorý bližšie rozoberieme v podkapitole 1.3. Pri riešení úlohy bol konkrétne použitý model párovej diferenciálnej Siamskej siete (Pairwise differential siamese network-PDSN). Tento model sa učí zhody medzi zakrytými časťami tváre a príznakovými prvkami poškodenými oklúziou. Potom sa podľa toho vytvorí slovník masiek, ktorý sa používa na zostavenie masky na vyradenie prvkov (FDM) pre

testovaciu tvár s náhodnými čiastočnými oklúziami.

Model Siamskej neurónovej siete ako metóda na riešenie problému opätovnej identifikácie osoby je bližšie spomenutá autormi v článku *Gated siamese convolutional neural network architecture for human re-identification*[12]. Základná idea Siamskej neurónovej siete je naučiť sa približovať podobné páry obrázkov, t.j. patriace jednej konkrétnej osobe a naopak vzdďaľovať od seba odlišné páry obrázkov, t.j. obrázky patriace dvom rôznym osobám. Ako z názvu článku vyplýva autori prišli s modelom *Gated siamese neural network*, to znamená, že model vychádzal zo Siamskej neurónovej siete a k nemu autori pridali gating funkciu na porovnanie extrahovaných príznakov párov obrázkov zo strednej vrstvy siete. Do vyšších vrstiev siete sa dostávali iba relevantné príznaky. Mechanizmy vo vnútri gating funkcie zvyšujú späťne šírené gradienty, ktoré zodpovedajú zosilneným miestnym podobnostiam medzi obrázkami. Toto podporuje nižšie a stredné vrstvy, aby sa učili filtre extrahovať iba príznaky, ktoré sú podobné a tým sa naučia odlišovať pozitívne páry od negatívnych párov. Táto metóda je pomenovaná "Matching Gate"(MG).

Autori v článku *A two stream sia-mese convolutional neural network for person re-identification*[2] navrhli ďalší model už vyššie spomenutej Siamskej neurónovej siete a to dvoj-streamovú(two streams) siamskú konvolučnú neurónovú sieť. Každý stream je siamská neurónová sieť. Takáto architektúra umožňuje učenie priestorových a časových informácií sperátne. Autori navrhli váhovanú dvoj-streamovú trénovaciu objektovú funkciu, ktorá kombinuje cenu Siamskej siete na priestorové a časové streamy s cieľom prepovedať identitu osoby. Presnosť metódy bola demonštrovaná na datasetoch PRID2011 a iLIDS-VID, ktoré sú verejne prístupné. Množstvo tradičných prístupov na opätovnú identifikáciu osoby je založených na low-level príznakoch vo forme histogramov farieb a textur, ktoré využívajú metriku učenia na nájdenie

funkcie vzdialenosti medzi obrázkami z rovnakej triedy. To znamená, že obrázky, ktoré patria do rovnakej triedy vzdialenosť minimalizujú a obrázky, ktoré patria do odlišnej triedy vzdialenosť maximalizujú. Klasifikačné/učiace metódy navrhnuté autormi na opätovnú identifikáciu osoby používajú vyššie spomenutú učenie vzdialenosti na nájdenie funkcie vzdialenosti medzi obrázkami. Model siamskej neurónovej siete obsahuje dve identické podsiete, ktoré zdieľajú váhy a sú vhodné na nájdenie podobností medzi dvoma porovnateľnými vstupmi. V navrhovanom modeli sú vstupom do prvej siete dve sekvencie RGB framov, kde každá sekvencia je zachytená odlišnou kamerou. Druhá sieť spracováva optický tok informácií z oboch kamier. Architektúra oboch sietí je rovnaká. Autori nazvali sieť asociovanú s priestorovým obsahom SpatialNet a sieť asociovanú s časovým obsahom TemporalNet.

1.2 Databázy

V tejto podkapitole predstavíme často používané databázy v oblasti počítačového videnia, ktoré slúžia na rozpoznávanie tvári alebo na opätovnú identifikáciu konkrétnych osôb. Databázy slúžili pri testovaní veľkého množstva problémov v oblasti detekcie a rozpoznávania tvári. Niektoré datasety boli použité aj na detekciu oklúzie tváre, a tiež na sledovanie osôb. V jednotlivých podkapitolách uvedieme pre čitateľa zaujímavé databázy obrázkov tvári, stručne opíšeme rozsah, použitie jednotlivých databáz a ukážku obrázkov nachádzajúcich sa v databáze. Nižšie spomenuté databázy sú vo väčšine prípadov verejne prístupné, niektoré je možné použiť iba na akademické účely. V článku Face databases[5] sa nachádzajú podrobnejšie vysvetlené databázy tvári.

1.2.1 Labeled faces in the wild

Databáza Labeled faces in the Wild určená pre štúdium rozpoznávania tváre bola uvedená v roku 2007[6]. Databáza obsahuje má nasledovný rozsah:

- 13233 obrázkov tvárí
- 5749 osôb
- 1680 osôb má viac než jeden obrázok
- 4069 osôb má iba jeden obrázok

Obrázky majú formát JPEG s rozmerom 250x250, pričom väčšina obrázkov je farebných, a malé množstvo je v odtieňoch šedej. Na obrázkoch ľudí je možné vidieť rôzne variácie. V nasledovnom zozname uvedieme niektoré variácie:

- pózy
- svetelné podmienky
- výrazy tváre
- prostredie
- rasa
- etnicita
- vek
- pohlavie
- oblečenie
- účes

- kvalita kamery
- farebná saturácia
- zaostrenie

Tváre v databáze boli detekované algoritmom Viola-Jones face detector[14]. Pri niektorých osobách je možné nájsť obrázky, ktoré obsahujú oklúziu tváre. Na obrázku 1.1 vidíme ukážku databázy Labeled faces in the Wild[6].



Obr. 1.1: Labeled faces in the Wild[6]

1.2.2 Aleix and Robert

Databáza obrázkov Aleix and Robert bola predstavená v roku 1998[8]. Databáza pozostáva z 3000 obrázkov. Osoby na obrázku sú zobrazené z predného pohľadu. Na obrázkoch je možné vidieť rôzne výrazy tváre ako napríklad hnev, úsmev, neutrálny výraz, ďalej svetelné podmienky a rôzne druhy oklú-

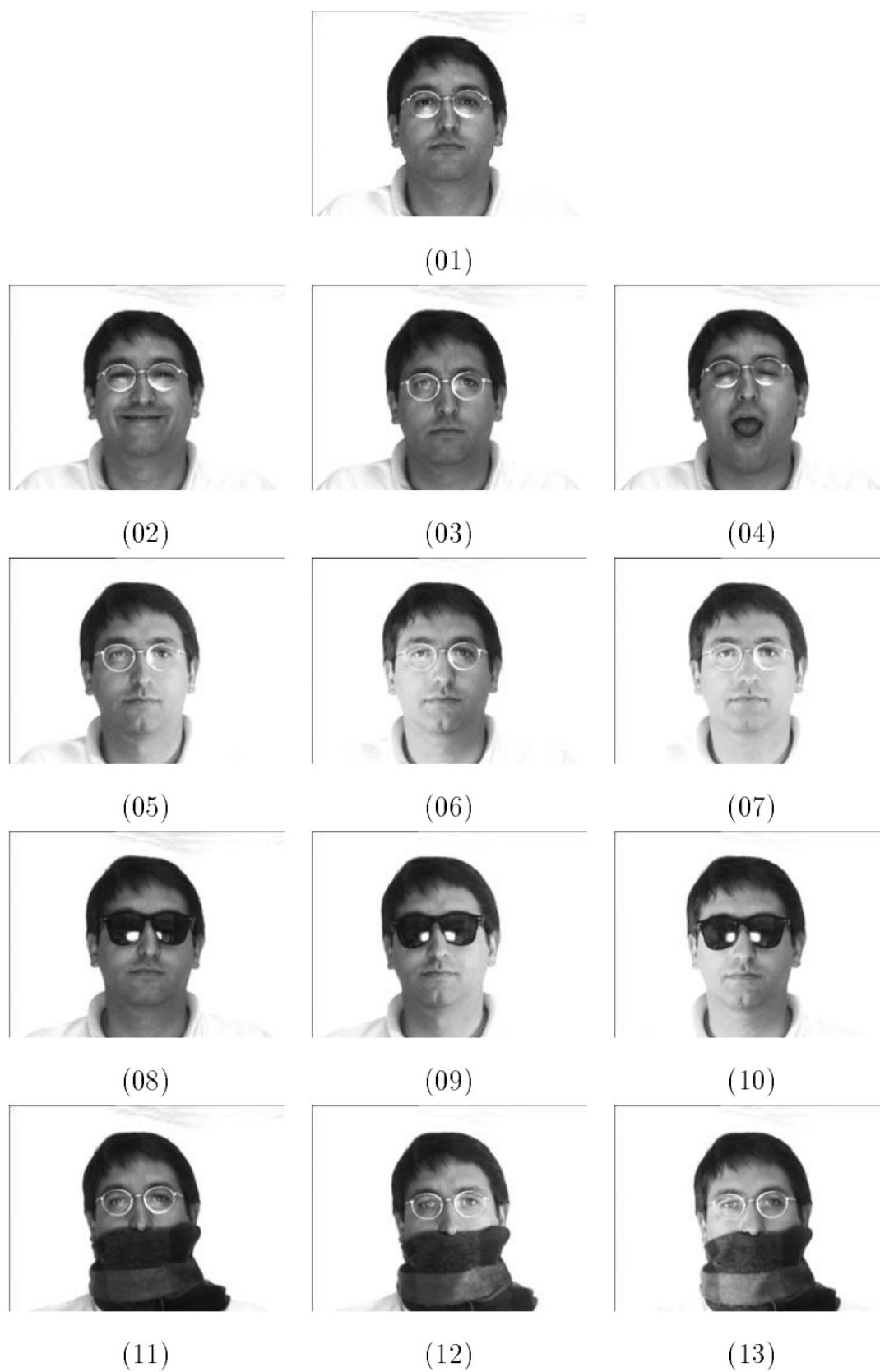
zí tváre, napríklad tvár zakrytá slnečnými okuliarmi alebo šatkou. Databáza má nasledovný rozsah:

- 116 osôb zobrazovaných na celkovo 3000 obrázkoch
- 63 obrázkov mužov
- 53 obrázkov žien
- každá osoba má 26 obrázkov

Osoby na obrázkoch mali povolené mať oblečené akékoľvek oblečenie a tiež neboli bližšie špecifikované podmienky na make-up, účes a iné. Obrázky boli nafotené z rovnakej vzdialenosti a systém na fotografovanie bol kalibrovaný dva krát denne. V nasledovnom zozname uvedieme zaujímavé parametre použitého systému:

- Pentium 133MHz, 64MB RAM, 2Gb HD
- Typ farebnej kamery: SONY 3CCDs
- 12 mm optika
- Frame grabber: Matrox Meteor RGB

Ukážku obrázkov z datasetu Aleix and Robert môžeme vidieť na obrázku 1.2. Čitateľ si môže všimnúť jednotlivé variácie obrázkov, ktoré obsahujú oklúziu tváre, sú rôzne osvetlené a osoba mení výrazy tváre.



Obr. 1.2: Aleix and Robert databáza[8]

Napriek tomu, že databáza vznikla pomerne dávno v súčasnosti patrí stále k veľmi používaným v oblasti rozpoznávania tváre.

1.2.3 Face occlusion dataset

Databáza Face occlusion dataset bola predstavená v článku Face occlusion detection using deep convolutional neural network[15]. Táto databáza nie je verejne prístupná. Autori zostavili túto databázu pri riešení problému detegovania oklúzie tváre, tiež boli použité vyššie spomenuté databázy Labeled faces in the Wild[6] a Aleix and Robert[8]. Geometrické a textové informácie z komponentov tvare - oči, nos a ústa su dôležité pri rozpoznávaní tváre , ale aj pri opätovnej identifikácii konkrétnej osoby. Ak je niektorá z týchto častí zakrytá, považujeme to za oklúziu. Za bežné oklúzie sú považované oklúzie klobúkom, slenčnými okliarmi, maskou, prilbou alebo akýmkoľvek iným predmetom. Bežné oklúzie tváre môžeme vidieť na obrázku 1.3.



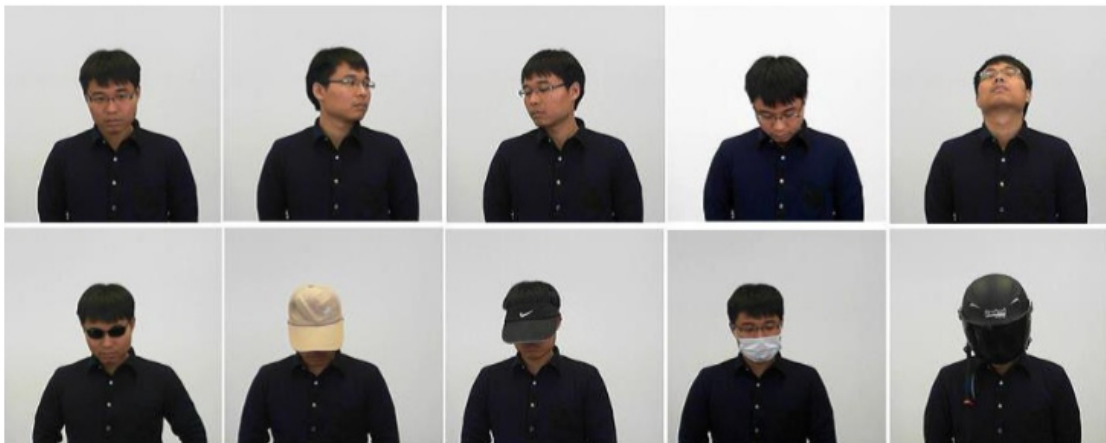
Obr. 1.3: Rôzne oklúzie tváre[15]

Rozsah databázy Face occlusion dataset je zhrnutý v nasledovnom zozname:

- 220 osôb

- 140 obrázkov mužov
- 80 obrázkov žien

Subjekty boli nafotené s rôznymi typmi oklúzie, tiež natáčali hlavu do rôznych smerov pod 45 stupňovým uhlom. Podobne ako pri zostavovaní databázy Aleix and Robert ani v tomto prípade neboli obmedzenia na druh oblečenia, výrazu tváre alebo spôsobu líčenia tváre. Ukážku obrázkov z datasetu Face occlusion dataset môžeme vidieť na obrázku 1.4.



Obr. 1.4: Rôzne oklúzie tváre[15]

Pre účely riešenia problému detekcie oklúzie tváre [15], ktoré sme spomínali vyššie v podkapitole 1.1 Suvisiace práce, autori zo sady obrázkov hornej polovice tela osoby detekovali tvár. Ukážku zdetegovaných tvári osôb je zobrazená na obrázku 1.5



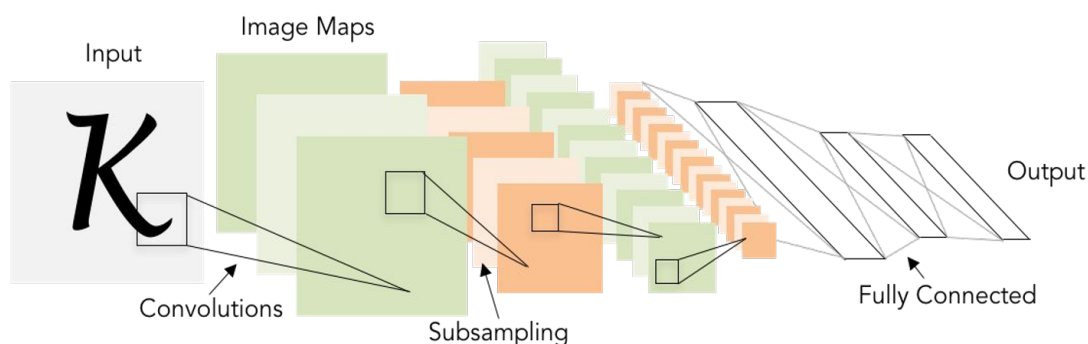
Obr. 1.5: Rôzne oklúzie tváre[15]

1.3 Konvolučné neurónové siete

Konvolučná neurónová sieť[3] je typ hlbkej doprednej neurónovej siete , ktorá sa používa na analýzu obrazových dát. Každý obraz je reprezentovaný viacrozmerným vektorom. Rozmery vektoru sú šírka, výška a hĺbka. Konvolučná neurónová sieť je sekvencia vrstiev a každá vrstva transformuje objem aktivácií do iného prostredníctvom diferencovateľnej funkcie. Architektúra konvulčnej neurónovej siete pozostáva z nasledovných vrstiev:

- Konvolučná vrstva
- Pooling vrstva
- Plne prepojená vrstva

Na obrázku 1.6 vidíme usporiadanie základného modelu konvolučnej neurónovej siete. Na vstupe neurónová sieť dostane obrázok, ktorý je postupne spracovaný konvolučnou vrstvou, pooling vrstvou a plne prepojenou vrstvou. Vstupná vrstva udržiava hodnoty pixelov obrázka vo formáte raw. Obrázky vo formáte raw nie sú komprimované ani inak upravované. Ako bolo spomenuté vstupný obrázok ma tri dimenzie a to šírku, výšku a hĺbku s tromi farebnými kanálmi R, G a B.



Obr. 1.6: Architektúra konvolučnej neurónovej siete

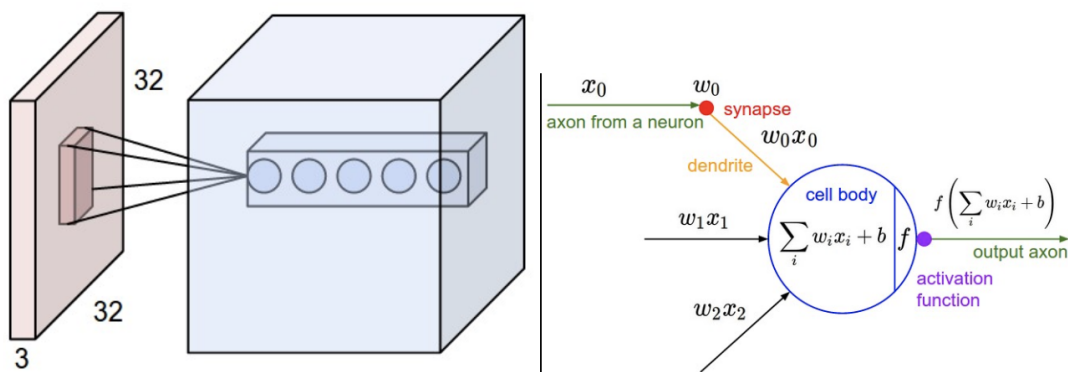
1.3.1 Konvolučná vrstva

Z názvu konvolučnej neurónovej siete vyplýva, že neurónová sieť vykonáva matematickú operáciu zvanú konvolúcia. Konvolúcia je základný matematický nástroj pre lineárnu filtráciu. Filtráciu konvolúciou používame vtedy, keď potrebujeme obraz vyhladiť, rozmazať hrany, odstrániť šum, prípadne ak chceme získať hranový obraz. Konvolučná vrstva je základom konvolučnej neurónovej siete a vykonáva väčšinu náročných výpočtov. Parametrami konvolučnej vrstvy sú filtre, ktoré je možné sa naučiť, ďalej počet filtrov, veľkosť stride-u a množstvo paddingu. Filter má určitú šírku a výšku, pričom hĺbka je rovnaká ako hĺbka vstupného obrazu. Na získanie príznakov vstupného ob-

razu sa filter posúva jednotlivo po pixeloch s určitým krokom tzv. stride-om. To znamená ak je stride rovný 1, filter sa posúva po jednom pixeli. Filter je možné posúvať aj s väčším stridom napríklad so stride-om 2 a vtedy sa filter posunie naraz o 2 pixely. Stride musí byť vhodne zadefinovaný, aby nepreteká cez vstupný obraz. Napríklad vstupný obraz o šírke 7 a výške 7 pri použití filtra 3x3 nie je vhodný stride 3. Veľkosť výstupného obrazu by nebolo celé číslo. Na overenie veľkosti výstupu pri použitej veľkosti stride-u slúži nasledovný vzťah:

$$(N - F)/stride + 1 \quad (1.1)$$

kde N je veľkosť vstupu a F veľkosť filtra. Veľkosť výstupu musí byť kladné celé číslo. Filter vždy rozširuje celú hĺbku vstupného objemu. Niekedy sa vstupnému objemu pridá padding, to znamená, že okraj sa rozšíri o nuly. Padding nám umožňuje kontrolovať priestorovú veľkosť výstupných objemov.



Obr. 1.7: Konvolučná vrstva[3]

1.3.2 Pooling vrstva

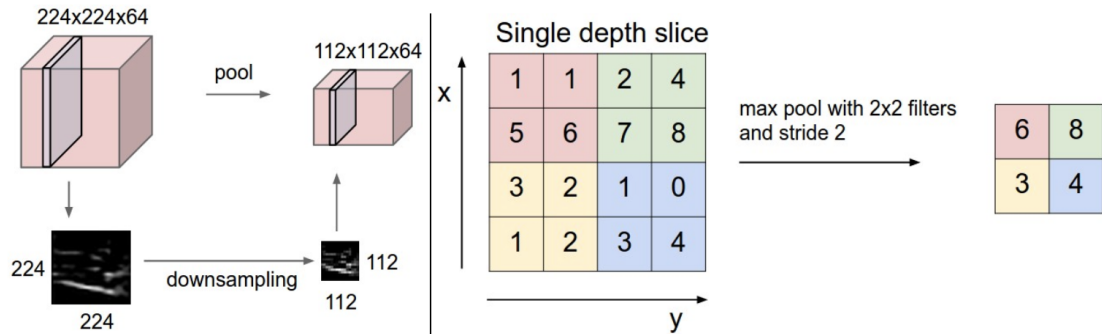
V architektúre modelu konvolučnej neurónovej siete sa pooling vrstva nachádza medzi konvolučnými vrstvami. Úlohou pooling vrstvy je znižovanie priestorovej veľkosti reprezentácie, aby sa znížilo množstvo parametrov a výpočtov v sieti, a tým sa zabránilo preučeniu siete. Pooling vrstva je definovaná nasledovnými parametrami:

- Objem o veľkosti $W_1 \times H_1 \times D_1$
- Priestorovy rozmer filtra F
- stride S

Pooling vrstva produkuje objem o veľkosti $W_2 \times H_2 \times D_2$, kde platí

- $W_2 = (W_1 - F) / S + 1$
- $H_2 = (H_1 - F) / S + 1$
- $D_2 = D_1$

V pooling vrstve sa na vstupnom objeme nepoužíva žiadny padding. Na obrázku 1.8 vľavo vidíme ako zmenší vstupný objem o veľkosti $224 \times 224 \times 64$ na výstupný objem o veľkosti $112 \times 112 \times 64$ použitím filtra s veľkosťou 2×2 so stride-om o veľkosti 2. Hĺbka výstupného objemu zostáva zachovaná. Najčastejšie používanou variáciou pooling vrstvy je txv. MaxPooling. Na obrázku 1.8 vpravo je znázornené akým spôsobom znižuje MaxPooling priestorovú veľkosť reprezentácie. Na vstupný objem je aplikovaný filter o veľkosti 2×2 s veľkosťou stride-u 2. Z každej časti vstupného objemu sa vyberie maximálna hodnota.



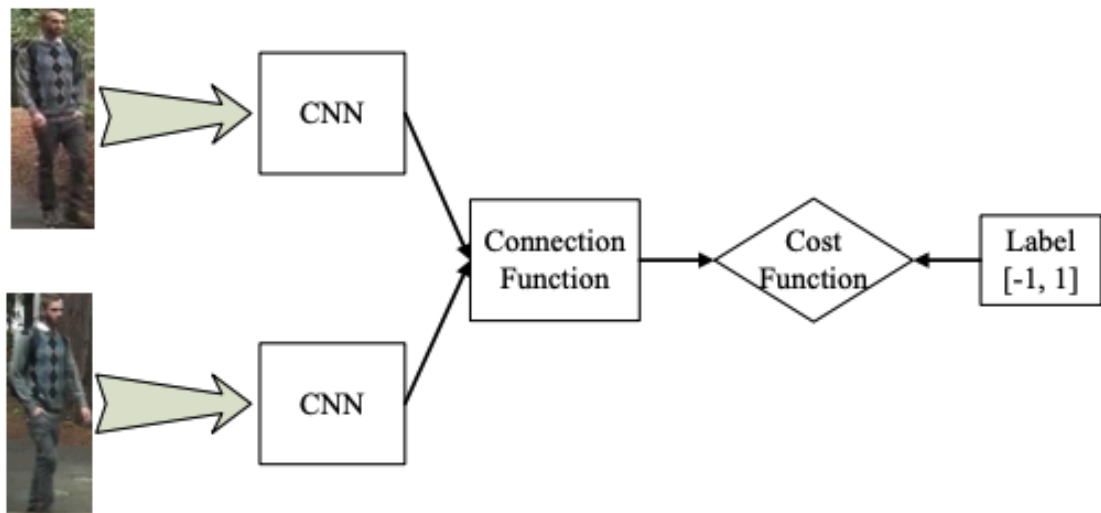
Obr. 1.8: Pooling vrstva[3]

1.3.3 Plne prepojená vrstva

Poslednou vrstvou v modeli konvolučnej neurónovej siete je plne prepojená vrstva. Plne prepojená vrstva obsahuje neuróny, ktoré sa pripájajú k celému vstupnému objemu ako v bežných neurónových sieťach. Plne prepojená vrstva agreguje informácie z konvolučnej vrstvy a pooling vrstvy, na základe ktorých je schopna klasifikovať vstupný obrázok do správnej triedy.

1.4 Siamská neurónová sieť

Siamská neurónová sieť pozostáva z dvoch identických sietí(dvojičiek), ktoré majú odlišné vstupy. Podsiete siamskej neurónovej siete zdieľajú rovnaké parametre, t.j. váhy a biasy. Najznámejšou aplikáciou siamskej siete je rozpoznávanie tváre, kde sa porovnávajú podobnosti tváre. Učenie siamskej siete je realizované pomocou triplet loss a contrastive loss.



Obr. 1.9: Siamska neurónová sieť

Na obrázku 4. je zobazený model siamskej neurónovej siete, ktorú možno rozdeliť na tri základné časti:

- dve podsiete
- connection funkcia
- cost funkcia

Connection funkcia sa používa na vyhodnotenie vzťahu medzi dvoma vstupmi a cost funkcia slúži na ohodnotenie vzťahu medzi dvoma vstupmi. Siamská sieť teda neslúži na klasifikáciu vstupu, ale hľadá podobnosti medzi dvoma vstupmi. Ak sú vstupy siete podobné, tak ich výstupné vektory sa budú nachádzať v priestore blízko seba. Výstupné vektory odlišných vstupov alebo málo podobných vstupov sa v priestore budú nachádzať ďalej od seba.

1.4.1 Triplet loss

1.5 Facenet

Model konvolučnej neurónovej siete Facenet bol predstavený v článku...

Kapitola 2

Návrh riešenia

Kapitola 3

Implementácia

Kapitola 4

Výsledky

Záver

Sem neskôr pribude záver.

Literatúra

- [1] AZEEM, A., SHARIF, M., RAZA, M., AND MURTAZA, M. A survey: Face recognition techniques under partial occlusion. *The International Arab Journal of Information Technology* 11 (01 2014), 1–10.
- [2] CHUNG, D., TAHBOUB, K., AND DELP, E. J. A two stream siamese convolutional neural network for person re-identification. In *2017 IEEE International Conference on Computer Vision (ICCV)* (Oct 2017), pp. 1992–2000.
- [3] CS231N. Convolutional neural network for visual recognition.
- [4] FARFADE, S. S., SABERIAN, M. J., AND LI, L. Multi-view face detection using deep convolutional neural networks. *CoRR abs/1502.02766* (2015).
- [5] GROSS, R. Face databases. In *Handbook of Face Recognition*, A. S. Li, Ed. Springer, New York, February 2005.
- [6] HUANG, G. B., RAMESH, M., BERG, T., AND LEARNED-MILLER, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

- [7] KURITA, T., PIC, M., AND TAKAHASHI, T. Recognition and detection of occluded faces by a neural network classifier with recursive data reconstruction. In *AVSS* (2003).
- [8] MARTINEZ, A. M. The ar face database. *CVC Technical Report* 24 (1998).
- [9] MATIYALI, N., AND SHARMA, G. Video person re-identification using learned clip similarity aggregation. *ArXiv abs/1910.08055* (2019).
- [10] SCHROFF, F., KALENICHENKO, D., AND PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. *CoRR abs/1503.03832* (2015).
- [11] SONG, L., GONG, D., LI, Z., LIU, C., AND LIU, W. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In *The IEEE International Conference on Computer Vision (ICCV)* (October 2019).
- [12] VARIOR, R. R., HALOI, M., AND WANG, G. Gated siamese convolutional neural network architecture for human re-identification. *CoRR abs/1607.08378* (2016).
- [13] VARIOR, R. R., HALOI, M., AND WANG, G. Gated siamese convolutional neural network architecture for human re-identification. In *Computer Vision – ECCV 2016* (Cham, 2016), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Springer International Publishing, pp. 791–808.
- [14] VIOLA, P., AND JONES, M. Robust real-time object detection. In *International Journal of Computer Vision* (2001).

- [15] XIA, Y., ZHANG, B., AND COENEN, F. Face occlusion detection using deep convolutional neural networks. *IJPRAI* 30 (2016), 1–24.
- [16] YI, D., LEI, Z., LIAO, S., AND LI, S. Z. Deep metric learning for person re-identification. In *2014 22nd International Conference on Pattern Recognition* (Aug 2014), pp. 34–39.