



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV INTELIGENTNÍCH SYSTÉMŮ**

DEPARTMENT OF INTELLIGENT SYSTEMS

**ALGORITMUS PRE HODNOTENIE KVALITY  
SNÍMKY TVÁRE**

ALGORITHM FOR FACIAL IMAGE QUALITY ESTIMATION

**BAKALÁŘSKÁ PRÁCE**

BACHELOR'S THESIS

**AUTOR PRÁCE**

AUTHOR

**TOMÁŠ HUSÁR**

**VEDOUcí PRÁCE**

SUPERVISOR

**Ing. TOMÁŠ GOLDMANN**

BRNO 2024

## Zadání bakalářské práce



156094

Ústav: Ústav inteligentních systémů (UITS)  
Student: **Husár Tomáš**  
Program: Informační technologie  
Název: **Algoritmus pro hodnocení kvality snímku obličeje**  
Kategorie: Umělá inteligence  
Akademický rok: 2023/24

### Zadání:

1. Seznamte se s problematikou rozpoznávání osob podle obličeje a zjistěte, co nejvíce ovlivňuje přesnost strojového rozpoznávání.
2. Na základě poznatků z předchozího bodu vytvořte dataset s kvalitními a nekvalitními snímky obličeje, který je určený pro trénování algoritmů strojového učení.
3. Sumarizujte informace o možnostech vyhodnocování kvality snímků obličeje pomocí neuronových sítí.
4. Navrhněte neuronovou síť a jednoduchou aplikaci, která bude provádět vyhodnocování kvality snímků obličeje.
5. Navržené řešení implementujte v programovacím jazyce Python.
6. Proveďte experimenty zaměřené na určení spolehlivosti vyhodnocení kvality.

### Literatura:

- SANG, Jitao; LEI, Zhen; LI, Stan Z. Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5. In: *Advances in Biometrics: Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings 3*. Springer Berlin Heidelberg, 2009. p. 229-238.
- KARAHAN, Samil, et al. How image degradations affect deep cnn-based face recognition?. In: *2016 international conference of the biometrics special interest group (BIOSIG)*. IEEE, 2016. p. 1-5.

Při obhajobě semestrální části projektu je požadováno:  
Body 1,2 a 3.

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

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

## Abstrakt

Presnosť algoritmov rozpoznávania tváre je výrazne ovplyvňovaná kvalitou vstupných obrázkov. Cieľom práce je vyhodnotiť kvalitu obrázkov tváří použitím konvolučnej neurónovej siete. Dáta, na ktorých bolo uskutočnené tréningovanie vznikli rôznymi degradáciami fotiek z CelebA datasetu. Výsledná aplikácia určuje kvalitu snímok na základe predikovaných pravdepodobností jednotlivých degradácií.

## Abstract

The precision of face recognition algorithms is heavily influenced by the quality of input images. The aim of the work is to evaluate the quality of face images using a convolutional neural network. The data on which the testing was carried out were created by various degradations of photos from the CelebA dataset. The resulting application determines the quality of images based on the predicted probabilities of individual degradations.

## Klíčové slová

počítačové videnie, rozpoznávanie tváre, hodnotenie kvality obrazu tváre, konvolučná neurónová sieť, ResNet

## Keywords

computer vision, face recognition, face image quality estimation, convolutional neural network, ResNet

## Citácia

HUSÁR, Tomáš. *Algoritmus pre hodnotenie kvality snímky tváre*. Brno, 2024. Bakalárska práca. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Tomáš Goldmann

# Algoritmus pre hodnotenie kvality snímky tváre

## Prehlásenie

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

.....  
Tomáš Husár  
9. mája 2024

## Podakovanie

Ďakujem vedúcemu práce Ing. Tomášovi Goldmannovi, ktorý mi poskytol odbornú pomoc  
.



# Obsah

<b>1</b>	<b>Úvod</b>	<b>4</b>
<b>2</b>	<b>Rozpoznávanie osôb podľa tváre</b>	<b>5</b>
2.1	Zariadenia pre snímanie tváre . . . . .	6
2.2	Strojové učenie a neurónové siete . . . . .	7
2.2.1	Nekontrolované učenie . . . . .	7
2.2.2	Kontrolované učenie . . . . .	8
2.3	Algoritmy pre detekciu tváre . . . . .	16
2.4	Algoritmy pre rozpoznávanie tváre . . . . .	20
2.5	Etické a právne aspekty rozpoznávania tváre . . . . .	25
<b>3</b>	<b>Hodnotenie kvality snímky</b>	<b>26</b>
3.1	Možnosti vyhodnocovania kvality . . . . .	26
3.1.1	HVS orientované IQA metriky . . . . .	26
3.1.2	IQA metriky pri rozpoznávaní tváre . . . . .	28
3.2	Faktory ovplyvňujúce rozpoznávanie tváre . . . . .	29
3.2.1	Kontrolované snímanie . . . . .	30
3.2.2	Snímanie v reálnych podmienkach . . . . .	31
<b>4</b>	<b>Návrh a implementácia</b>	<b>32</b>
4.1	Príprava datasetu . . . . .	32
4.1.1	Degradácia snímok . . . . .	33
4.1.2	Rozdelenie snímok v datasete . . . . .	40
4.1.3	Predspracovanie snímok . . . . .	42
4.2	Implementácia architektúry ResNet . . . . .	43
4.3	Výsledná aplikácia . . . . .	44
4.4	Skóre kvality . . . . .	44
<b>5</b>	<b>Experimenty</b>	<b>46</b>
5.1	Hodnotiace metriky . . . . .	48
5.2	Zmeny v datasete . . . . .	51
<b>6</b>	<b>Záver</b>	<b>53</b>
	<b>Literatúra</b>	<b>54</b>
<b>A</b>	<b>Obsah SD</b>	<b>60</b>

# Zoznam obrázkov

2.1	Tok spracovania dát pri rozpoznávaní tváre. . . . .	5
2.2	Klastrovanie pomocou <i>k-means</i> algoritmu [3]. . . . .	8
2.3	Klasifikácia medzi dvomi triedami pomocou SVM. . . . .	9
2.4	Fungovanie algoritmu AdaBoost. . . . .	10
2.5	Štruktúra jednoduchkej neurónovej siete. . . . .	11
2.6	Reziduálne bloky architektúry ResNet. . . . .	14
2.7	Príklad HOG z obrázka . . . . .	16
2.8	Príklady prvých Haar Príznakov pri Viola-Jones . . . . .	17
2.9	Hľadanie sumy obdĺžnikovej oblasti . . . . .	17
2.10	Kaskáda slabých klasifikátorov algoritmu Viola-Jones . . . . .	18
2.11	Detekcia objektov pomocou SSD algoritmu [42]. . . . .	18
2.12	Detekcia objektov pomocou YOLO algoritmu [54]. . . . .	19
2.13	Architektúra RetinaFace [11]. . . . .	20
2.14	Porovnanie PCA a LDA projekcií [48] . . . . .	21
2.15	Rozloženie vzoriek tváří pomocou algoritmu SphereFace [44]. . . . .	22
2.16	Rozdiely pri použití stratovej funkcie Softmax a ArcFace [12]. . . . .	23
2.17	Rozdiely metód Softmax, ArcFace a SphereFace pri binárnej klasifikácii[12].	23
2.18	Magnitúdy rôznych snímok tváre vzhľadom na stred triedy pri MagFace [47].	24
3.1	Hodnotenie kvality snímky tváre pomocou FaceQNet [30]. . . . .	28
3.2	Vyhodnotenie asymetrie tváre a svetelných podmienok pomocou Gáborových vlniek [55]. . . . .	29
3.3	Transformácia obrázkov do frekvenčnej domény pomocou DCT [55]. . . . .	29
3.4	Štandardné parametre snímky tváre v súlade s ISO/IEC 19794-5:2005[34]. .	30
4.1	Diagram navrhutej aplikácie. . . . .	32
4.2	Snímka bez poškodení. . . . .	33
4.3	Graf generického gaussovho pravdepodobnostného rozdelenia. . . . .	33
4.4	Diskrétna aproximácia gaussovho jadra o veľkosti $3 \times 3$ a $5 \times 5$ [62]. . . . .	34
4.5	Obrázok poškodený rozmazaním. . . . .	35
4.6	Obrázok poškodený šumom. . . . .	36
4.7	Obrázok poškodený kompresiou. . . . .	37
4.8	Hraničné body hornej časti tváre, a oblasti, ktoré ohraničujú. . . . .	38
4.9	Obrázok poškodený zakrytím hornej časti tváre. . . . .	39
4.10	Hraničné body spodnej časti tváre, a oblasti, ktoré ohraničujú. . . . .	40
4.11	Obrázok poškodený zakrytím spodnej časti tváre. . . . .	40
4.12	Rozdelenie snímok v FIQVD. . . . .	41
4.13	Spracovanie obrázka na vstup do modelu. . . . .	42
4.14	Predikcia pomocou stratovej funkcie Cross-Entropy Loss. . . . .	43

4.15	Predikcia pomocou stratovej funkcie MSE. . . . .	43
4.16	Snímka výslednej aplikácie. . . . .	44
5.1	Logaritmizované stratové hodnoty funkcie Cross-Entropy-Loss modelu ResNet-50. . . . .	46
5.2	Logaritmizované Stratové hodnoty funkcie MSE modelu ResNet-50. . . . .	47
5.3	Stratové hodnoty funkcie Cross-Entropy-Loss modelu ResNet-18. . . . .	47
5.4	Stratové hodnoty funkcie MSE modelu ResNet-18. . . . .	48
5.5	Rozdelenie vzoriek podľa ich predikcie a skutočného označenia [3]. . . . .	48

# Kapitola 1

## Úvod

Technológia rozpoznávania tváre nadobudla široké uplatnenie v rôznych sférach vďaka svojej schopnosti presnej a spoľahlivej identifikácie jednotlivcov na základe jedinečných charakteristík ich tváre. Danú technológiu je možné použiť nielen pri úlohách autentifikácie ako napríklad odomykanie osobných zariadení a potvrdzovanie finančných transakcií, ale aj pri identifikácii podozrivých alebo nezvestných osôb zachytených kamerovými systémami. Tieto systémy slúžia k monitorovaniu verejného priestranstva a sú súčasťou každého rozvinutého mesta.

V kontexte rozpoznávania tváre je presnosť týchto systémov priamo ovplyvnená kvalitou vstupných snímok. Nekvalitné snímky sú snímky, na ktorých je náročné (alebo takmer nemožné) extrahovať kľúčové črty tváre snímanej osoby, na ktorej sa rozpoznávanie vykonáva. Ide o rozmazané snímky, snímky obsahujúce šum či kompresné artefakty, alebo snímky, na ktorých je obmedzená viditeľnosť kľúčových črt tváre z dôvodu nízkeho rozlíšenia kamery, nerovnomerného osvetlenia tváre, rotácie hlavy alebo oklúzie spôsobenej objektmi ako sú slnečné okuliare, masky, pokrívky hlavy a pod.

Rozmazanie snímky je spôsobené nedostatočným zaostrením snímacieho zariadenia na tvár jednotlivca alebo pohybom snímacieho zariadenia či snímaného objektu. Šum môže byť spôsobený náhodnými signálmi v elektronických komponentoch, alebo zvýšenou citlivosťou kamery (ISO), ktorá nastáva pri snímaní v slabom osvetlení. Kompresné artefakty vznikajú pri kompresii obrazových dát za účelom prenosu cez sieť, alebo ušetrenia pamäti potrebnej k ich uloženiu.

Schopnosť vyhodnotenia kvality snímky je dôležitým nástrojom pri účeloch rozpoznávania tváre, pretože určením „vhodnosti“ snímky je možné vybrať najvhodnejšiu snímku zo sekvencie obrazových dát pre navýšenie presnosti a minimalizáciu výpočtovej záťaže rozpoznávacích systémov.

Revolúciu počítačového videnia predstavujú konvolučné neurónové siete, ktoré umožňujú robustné a efektívne spracovanie dát pri rôznych úlohách, ako napríklad detekcia, rozpoznávanie alebo extrakcia kľúčových črt tváre.

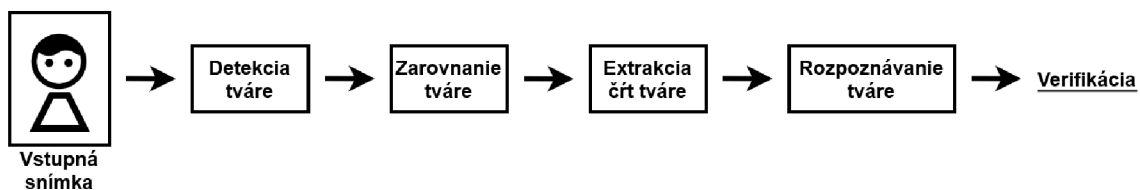
Aplikácia popísaná v tejto práci využíva konvolučnú neurónovú sieť architektúry ResNet, ktorá predikuje pravdepodobnosti jednotlivých typov poškodení snímky. Na základe týchto pravdepodobností a váh priradených daným typom poškodenia, ktoré boli určené na základe podobnosti pôvodných a poškodených snímok z CelebA datasetu, je vypočítané skóre, ktoré určuje kvalitu vstupnej snímky.

## Kapitola 2

# Rozpoznávanie osôb podľa tváre

Technológia rozpoznávania tváre je biometrická metóda, ktorá sa zakladá na identifikácii jednotlivcov na základe ich charakteristických črt tváre. Rozpoznávanie tváre je podproblémom v rámci širšieho kontextu rozpoznávania vzorov, kde je obraz reprezentovaný ako matica pixelov rôznych hodnôt, a algoritmy rozpoznávania tváre identifikujú jednotlivcov na základe črt tváre, ktoré sú zakódované v týchto dátach [60].

Rozpoznávanie osôb podľa tváre v širšom zmysle zahŕňa súvisiace technológie, vrátane detekcie tváre, určovania jej presnej pozície a zarovnanie tváre, extrakcie kľúčových črt a rozpoznávanie identity osoby. Postupnosť týchto operácií pri spracovaní snímky tváre je zobrazená na obrázku 2.1.



Obr. 2.1: Tok spracovania dát pri rozpoznávaní tváre.

Detekcia tváre predstavuje proces lokalizácie a identifikácie prítomnosti tvárí v rámci vstupného obrázka. Výstupom detekcie je zvyčajne ohraničená oblasť obsahujúca tvár vo forme štvorca alebo obdĺžnika.

Pri zarovnávaní tváre sa lokalizujú charakteristické body na identifikovanej tvári, ako napríklad stred očí, kútiky úst a špička nosa. Výstup zarovnanie obsahuje súradnice týchto charakteristických bodov alebo samotný obrázok, ktorý bol zarovnaný (alebo inak normalizovaný) na základe týchto bodov.

Extrakcia relevantných črt zo zarovnaných oblastí má za účel zachytiť charakteristické vlastnosti tváre. Výstupom tohto procesu je *feature vector* (opísaný v sekcii 2.2) obsahujúci informácie o štruktúre tváre, ktoré možno použiť pri samotnom rozpoznávaní.

Rozpoznanie tváre je proces identifikácie jednotlivca, pri ktorom dochádza k porovnávaniu extrahovaných črt s hodnotami v databáze. Jeho výstupom je identita rozpoznaného jednotlivca.

Celkový úspech algoritmov rozpoznávania osôb závisí na kvalite extrakcie príznakov, presnosti porovnávanie a dostupnosti dostatočného množstva referenčných údajov na tréningovanie a overenie systému. V moderných systémoch sa často využívajú metódy hlbokého učenia a neurónové siete na zlepšenie výkonu a schopnosti adaptácie na rôzne podmienky.

## 2.1 Zariadenia pre snímanie tváre

Pri rozpoznávaní tváre sa využívajú rôzne typy snímacích zariadení a technológie, ktoré pomáhajú pri zachytávaní a analýze obrazových dát tváre. Pokroky v technológiách zaznamenávania umožňujú nové prístupy v analýze a spracovaní obrazu.

### **CCTV (*Closed Circuit Television*) kamery**

Sú tradičné záznamové zariadenia, ktoré sa používajú na monitorovanie rôznych verejných alebo súkromných priestorov. Termín „closed-circuit“ vychádza z faktu, že signály týchto kamier sú prenášané cez uzavretú sieť a dáta sa konvertujú napríklad na DVR formát. Podmienky pri rozpoznávaní tváre pomocou CCTV sú náročnejšie ako pri kontrolovanom rozpoznávaní, najmä kvôli rôznym podmienkam osvetlenia a pózy jednotlivcov [59].

### **IP kamery**

IP kamery sú v dnešnej dobe jedným z najpoužívanejších typov bezpečnostných snímacích zariadení. IP (Internet Protocol) kamery konvertujú videozáznamy do IP paketov, ktoré sú prenášané cez internetovú sieť. V závislosti od inštaláčnej konfigurácie (napríklad káblové pripojenie k interným sieťam alebo pripojenie kamery k riadiacemu systému) je možné IP kamery považovať za novodobý variant CCTV [21].

### **Kamery na osobné použitie**

Kamery na osobné použitie sú zväčša kamery pripojené k počítaču, laptopu alebo smartfónu. Sú navrhnuté pre individuálne využitie ich vlastníka. Vzhľadom na technologický pokrok posledných rokov môžu byť tieto kamery použité aj na účely rozpoznávania tváre, napríklad identifikáciou pri odomykaní telefónu alebo bankových aplikácií [64].

### **3D kamery**

Využívajú sa pri zachytávaní informácií o hĺbke obrazu spolu s vizuálnymi dátami. Tieto kamery môžu používať technológiu stereoskopického videnia, kde sa pri výpočte hĺbky spoliehajú na malé rozdiely medzi snímkami zachytenými viacerými šošovkami. Predstavujú pokročilý prístup k technológii rozpoznávania tváre, keďže zachytávajú informácie o trojrozmernom usporiadaní objektov. Tento prístup môže viesť k zlepšeniu presnosti a spoľahlivosti rozpoznávania tváre v porovnaní s tradičnými 2D kamerami [23].

### **LIDAR (*Light Detection and Ranging*) senzory**

Využívajú laserové svetlo na presné odmeranie vzdialenosti. Tieto senzory merajú čas, ktorý je potrebný na návrat laserového signálu od objektu, a na základe týchto dát sa vytvára detailná 3D mapa objektu alebo okolia. V praxi sú LIDAR senzory často implementované v odvetviach robotiky alebo senzorov autonómnych vozidiel. Aj keď niektoré osobné zariadenia, napríklad niektoré modely smartfónov Samsung a iPhone, majú tieto senzory, ich hlavné využitie spočíva v oblasti rozšírenej reality (AR) a meraní vzdialeností medzi bodmi. Napriek existencii navrhnutých riešení pre rozpoznávanie tváre pomocou týchto senzorov, ako je napríklad CloudNet [38], ich široké používanie v súčasnosti nie je bežné [46].

## 2.2 Strojové učenie a neurónové siete

Strojové učenie (ML) je odvetvie umelej inteligencie (AI), ktoré sa zameriava na vývoj algoritmov a techník. Hlavnou myšlienkou strojového učenia je umožniť počítačom automatizované učenie sa z dát a vytváranie modelov, ktoré dokážu vykonávať úlohy bez explicitného programovania.

K správne mu pochopeniu strojového učenia je nevyhnutné vysvetliť kľúčové pojmy spojené s ním. *Dataset*, alebo dátová sada, je súbor údajov pozostávajúci z viacerých dát (niekedy nazývaných ako vzorky). Každá vzorka, ktorú chceme analyzovať je predstavovaná svojím dátovým bodom. *Features* sú údaje, ktoré popisujú merateľné vlastnosti alebo charakteristiky jednotlivých vzoriek. Každá vzorka je reprezentovaná vo vysokorozmernom priestore, tzv. *Feature space*, ktorý je definovaný množinou *features*. Samotná reprezentácia vzorky v tomto priestore sa nazýva *Feature vector*. Tieto vektory pozostávajú z hodnôt zodpovedajúcich každej *feature* danej vzorky. K meraniu podobnosti (alebo odlišnosti) medzi vzorkami sa používajú vzdialenostné metriky. Najjednoduchšou vzdialenostnou metrikou je euklidovská vzdialenosť. Výpočet normalizovanej<sup>1</sup> euklidovskej vzdialenosti medzi dvomi bodmi môžeme definovať ako [3]:

$$dist(A, B) = \sqrt{\frac{\sum_{i=1}^n (a_i - b_i)^2}{n}} \quad (2.1)$$

Kde:  $n$  je počet *features*,  $a_i$  a  $b_i$  sú hodnoty jednotlivých *features* dátových bodov  $A$  a  $B$  a  $dist$  je vzdialenosť medzi týmito bodmi.

Avšak v závislosti od typu dát sa môžu využívať komplexnejšie metriky, ako kosínusová podobnosť *sim* merajúca kosínus uhlu medzi dvoma vektormi definovaná ako [32]:

$$sim(A, B) = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| |\vec{B}|} = \cos(\theta) \quad (2.2)$$

Kde: čitateľ predstavuje bodový súčin vektorov  $\vec{A}$  a  $\vec{B}$ , menovateľ je súčinom ich normalizovaných euklidovských dĺžok a  $\theta$  predstavuje uhol medzi vektormi  $\vec{A}$  a  $\vec{B}$ .

Na základe označenia dát pri úlohách strojového učenia môžeme rozdeliť do dvoch hlavných kategórií, a to nekontrolované učenie (*unsupervised learning*) a kontrolované učenie (*supervised learning*).

### 2.2.1 Nekontrolované učenie

Pri analýze údajov často nepoznáme skutočné označenia dát (*labels*) alebo chceme preskúmať prirodzene vznikajúce vzory v dátach. K tomuto účelu sa využívajú metódy nekontrolovaného strojového učenia (*unsupervised ML*), ako sú napríklad klastrovacie metódy<sup>2</sup> (*Clustering*) alebo redukcia dimenzionality dát.

Cieľom klastrovacích metód je identifikovať relevantné podskupiny v danej dátovej sade bez vopred definovanej hypotézy vlastností, ktoré by dané podskupiny mohli mať [3].

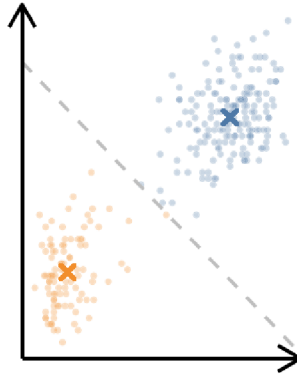
Príkladom klastrovacej metódy je *k-means clustering* [28], kde je počet klastrov, ktoré sa majú definovať preddefinovaný parametrom  $k$ . Každý klaster je reprezentovaný jeho stredom, čo je umelo vytvorený dátový bod predstavujúci strednú hodnotu všetkých bodov priradeným tomuto klastru.

<sup>1</sup>Normalizovaná euklidovská vzdialenosť  $dist$  medzi dvoma bodmi dosahuje hodnoty od 0 do 1, kde  $d = 0$  predstavuje identické (alebo takmer identické) vektory, a  $d = 1$  predstavuje navzájom ortogonálne vektory.

<sup>2</sup>Klaster je podmnožina dát, ktoré majú podobné vlastnosti alebo charakteristiky.

Na začiatku algoritmu sú stredy klastrov (*seeds*) náhodne umiestnené do *feature space*. Daný algoritmus sa následne iteruje cez dva kroky, kde sú dáta najprv priradené k najbližšiemu stredu klasteru, a následne sa aktualizuje pozícia všetkých stredov na základe kompozície vyplývajúcej z prvého kroku.

Po určitom počte iterácií zvyčajne nastane konvergencia k lokálnym optimám, kedy sa výrazne nemenia body, ktoré sú klastrom priradené (obr. 2.2).



Obr. 2.2: Klastrovanie pomocou *k-means* algoritmu [3].

Narastajúcim počtom *features*, a teda aj počtom rozmerov *feature space* sa značne komplikuje proces vizualizácie a analýzy dát [3].

Pre zmiernenie týchto problémov možno uplatniť metódy redukcie dimenzionality dát. Tieto metódy sa snažia zachovať čo najviac informácií z pôvodných dátových priestorov, ale zároveň znížiť ich rozmer na menej zložité podmnožiny. Jednou z najčastejších metód je *Principal Component Analysis* (PCA), ktorá identifikuje lineárne nezávislé zložky dát s najväčšou varianciou<sup>3</sup> a transformuje dáta do nového priestoru s menším počtom dimenzií. PCA je podrobnejšie popísané v sekcii 2.4.

### 2.2.2 Kontrolované učenie

Pri metódach kontrolovaného strojového učenia (*Supervised ML*) je cieľom algoritmov naučiť sa všeobecné pravidlá, ktoré mapujú vstupy na výstupy, aby bolo možné predikovať výstup pre nové dáta (iné dáta, ako počas procesu učenia), pri ktorých sú známe iba ich vstupné hodnoty. Existujú dve hlavné kategórie kontrolovaného učenia, a to klasifikácia, kde je výstupom zo vstupných dát kategória (tzv. trieda), do ktorej vstupné dáta patria, a regresia, kde sú výstupné hodnoty zo vstupných dát číselné [3].

Prelomovou metódou kontrolovaného učenia pri úlohách klasifikácie aj regresie sú neurónové siete (*neural networks* - NN) [25], ktoré sú schopné automaticky naučiť sa zložité vzťahy a reprezentácie v dátach. Tieto siete sa skladajú z viacerých vrstiev neurónov, ktoré spracujú vstupné dáta, a generujú výstupy na základe týchto dát. Neurónové siete sú podrobnejšie popísané nižšie v podsekcii 2.2.2.

Riešenie problémov viactriednych klasifikácií pomocou je možné stratégiami jeden proti všetkým (*One-vs-All* - OvA) alebo jeden proti jednému (*One-vs-One* - OvO)[26].

Pri stratégií OvA je trénovaný separátny binárny klasifikátor pre každú triedu. Počas trénovania sa vzorky patriace do aktuálnej triedy považujú za pozitívnu triedu a vzorky

<sup>3</sup>Variancia je štatistický pojem, ktorý vyjadruje mieru, akou sa dáta v dátovej množine líšia od svojho priemeru alebo očakávanej hodnoty.



zo všetkých ostatných tried za negatívnu. V čase predikcie každý klasifikátor vytvorí skóre a trieda s najvyšším skóre je označená ako predpovedaná trieda.

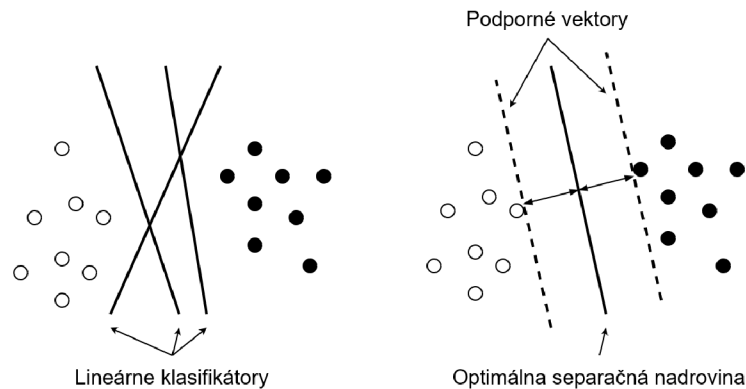
Pri stratégii OvO sa trénuje binárny klasifikátor pre každý pár tried. Počas tréovania je každý klasifikátor trévaný na podmnožine údajov obsahujúcej vzorky iba z tých dvoch tried, ktoré rozlišuje. V čase predikcie každý klasifikátor hlasuje za jednu z tried, a trieda s najvyšším počtom hlasov je označená za predpovedanú triedu.

OvA prístup je výpočetne efektívnejší pri problémoch s vysokým počtom tried, zatiaľ čo OvO býva odolnejší voči nevyváženosti údajov v datasetoch.

## Metódy podporných vektorov

Metóda podporných vektorov (*Support Vector Machines* - SVM) [66] je kontrolovaná metóda strojového učenia používaná na klasifikáciu dát predstavená v roku 1995. Mapuje vstupné vektory do vysokorozmerného priestoru cez lineárne alebo nelineárne mapovanie. V tomto priestore je skonštruovaná optimálna separačná nadrovina (podpriestor o jednu dimenziu menší ako pôvodný).

SVM bol vyvinutý na riešenie problému klasifikácie dvoch tried, kde cieľom je separácia týchto tried funkciou odvodenou z dostupných príkladov. V príklade na obrázku číslo 2.3 vľavo, môžeme vidieť veľa možných lineárnych klasifikátorov, ktoré separujú dáta. Avšak, existuje len jeden lineárny klasifikátor, zobrazený na obrázku 2.3 vpravo, ktorý maximalizuje vzdialenosť medzi nadrovinou a najbližšími bodmi každej triedy. Tento klasifikátor sa nazýva optimálna separačná nadrovina (Optimal Separating Hyperplane - OSH). Podporné vektory sú dátové body ležiace najbližšie k tejto nadrovine [27].



Obr. 2.3: Klasifikácia medzi dvomi triedami pomocou SVM.

V kontexte detekcie tvárí môže byť SVM trévaná na oblastiach obrázkov s tvármi a bez, pričom cieľom je naučiť klasifikátor rozlíšiť medzi týmito oblasťami.

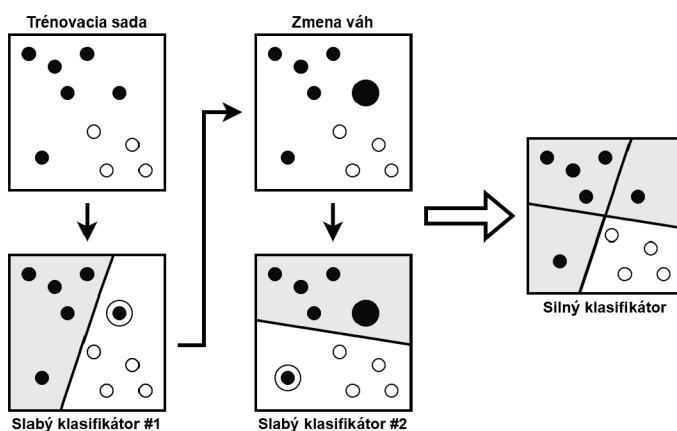
## AdaBoost

Algoritmus *Adaptive Boosting* (AdaBoost) [19], uvádzaný v roku 1995, predstavuje prvý praktický boosting algoritmus a zostáva jedným z najpoužívanejších dodnes. Boosting je prístup k strojovému učeniu, ktorý využíva myšlienku vytvorenia presnej predikcie kombináciou mnohých relatívne nepresných a slabých klasifikátorov. Jeho aplikáciu možno sledovať v mnohých oblastiach, vrátane počítačového videnia, spracovania prirodzeného jazyka, finančných analýz a bioinformatiky.

Slabý klasifikátor je model, ktorý funguje o niečo lepšie ako náhodný výber pri probléme binárnej klasifikácie. Môže to byť napríklad jednoduchý rozhodovací strom. AdaBoost iteratívne trénuje sekvenciu slabých klasifikátorov. V každej iterácii sa trénuje nový slabý klasifikátor, pričom sa kladie dôraz na inštancie, ktoré boli nesprávne klasifikované predchádzajúcimi klasifikátormi [56].

Každá inštancia v trénovacej sade má priradenú váhu, ktorá sa mení po každej iterácii. Inštanciám, ktoré boli nesprávne zatriedené, sa priradí vyššia váha, aby sa zabezpečilo, že nasledujúce klasifikátory budú viac zamerané na správne roztriedenie týchto inštancií.

Konečný silný klasifikátor je váženou kombináciou slabých klasifikátorov. Na základe presnosti každého klasifikátora počas trénovacieho procesu sa určuje jeho váha. Vo výslednom klasifikátore majú presnejšie klasifikátory vyššie váhy. Počas predikcie každý slabý klasifikátor hlasuje a konečná predikcia je určená na základe váhy jednotlivých hlasov. Tento proces je znázornený na obrázku 2.4.



Obr. 2.4: Fungovanie algoritmu AdaBoost.

## Neurónové siete

Neurónové siete (NN) [25] sú výpočtové modely inšpirované štruktúrou a fungovaním ľudského mozgu, ktoré nájdeme v rôznych oblastiach, ako je napríklad rozpoznávanie reči, tváří, odtlačkov prstov alebo dúhovky [37].

Pre správne pochopenie fungovania neurónových sietí je nevyhnutné oboznámiť sa s pojmami, ktoré úzko súvisia s touto technológiou.

Neurón, tiež známy ako uzol, predstavuje základnú výpočtovú jednotku. Každý neurón v sieti sa podieľa na spracovaní dát a celkovom výpočte danej siete. Prijíma vstupy, pričom každý vstup je spojený s parametrom váhy a hodnotou *bias*. Tieto parametre určujú vplyv neurónu na výstupné dáta. Matematicky je výstup neurónu  $y$  vyjadrený vzťahom[25]:

$$y = f\left(\sum_{i=1}^n (x_i \cdot w_i) + b\right) \quad (2.3)$$

Kde:  $x_i$  je vstup  $i$ -teho neurónu (alebo vonkajšieho zdroja),  $w_i$  je váha priradená  $i$ -temu vstupu,  $b$  je bias a  $f()$  je aktivačná funkcia.

Aktivačná funkcia je matematická funkcia, ktorá je používaná v neurónových sieťach na transformáciu váženej sumy vstupov neurónu na výstupnú hodnotu. Existuje niekoľko

rôznych typov aktivačných funkcií, z ktorých má každá svoje vlastnosti a opodstatnenie pri rôznych úlohách strojového učenia.

Jednou z najpoužívanejších aktivačných funkcií (najmä v skrytých vrstvách hlbokých sietí, ktoré sú vysvetlené o kus nižšie) vďaka svojej jednoduchosti a schopnosti rýchlej konvergencie je ReLU (*Rectified Linear Unit*) definovaná ako [49]:

$$f(x) = \max(0, x) \quad (2.4)$$

Kde: pre všetky záporné hodnoty vstupu  $x$  je  $f$  definovaná ako 0, a pre všetky kladné vstupy rastie lineárne, čo znamená, že ak je vstup kladný, výstup bude rovný vstupu. ReLU je veľmi jednoduchá funkcia, ktorá vykonáva nelineárne transformácie kľúčové pre správny proces učenia neurónových sietí.

Aktivačnou funkciou často používanou na konci neurónovej siete pri úlohách viactriednej klasifikácie, teda pri úlohách identifikácie kategórie alebo triedy, do ktorej patria vstupné dáta, je Softmax definovaný ako [49]:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \quad (2.5)$$

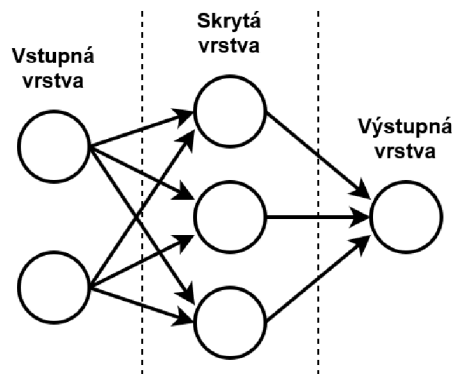
Kde  $x_i$  predstavuje element vstupného vektoru  $x$ . Softmax normalizuje výstup na rozsah 0 až 1, a zabezpečuje, aby všetky výstupy dávali pri súčte hodnotu 1. Výstup s najväčšou hodnotou je výsledná predikcia pre triedu, do ktorej by vzorka mala patriť.

Pri úlohách binárnej klasifikácie sa na konci neurónovej siete používa aktivačná funkcia Sigmoid, ktorá je definovaná ako [49]:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (2.6)$$

Daná funkcia transformuje vstup  $x$  na hodnoty v rozsahu 0 až 1, kde  $\text{Sigmoid}(x)$  vyjadruje pravdepodobnosť príslušnosti k jednej z dvoch tried.

Neuróny sú organizované do vrstiev a spojenia medzi nimi tvoria model neurónovej siete, ktorý predstavuje štruktúru, inak povedané architektúru, danej siete. Neuróny vo vstupnej vrstve predstavujú vstupné hodnoty spracovaných údajov. Neuróny v skrytej vrstve spracovávajú vstupné dáta prostredníctvom vážených spojení a aktivačných funkcií. Neuróny výstupnej vrstvy produkujú predikciu modelu, čiže konečný výstup siete (obr. 2.5).



Obr. 2.5: Štruktúra jednoduchého neurónovej siete.

Parametre modelu sa aktualizujú počas procesu tréningu, ktorým sa model učí z tréningových dát a aktualizuje svoje parametre pomocou optimalizačného algoritmu tak,

aby minimalizoval chybu predikcie na základe gradientu stratovej funkcie voči týmto parametrom. Zjednodušene povedané, stratová funkcia kvantifikuje rozdiely medzi predikovanými a cieľovými hodnotami (označenými pomocou tzv. *labels*) jednotlivých vzoriek, zatiaľ čo optimalizačný algoritmus určuje, akým spôsobom sa dané parametre menia. Proces aktualizácie parametrov je súčasťou algoritmu spätnej propagácie *backpropagation*, ktorý je základným mechanizmom učenia neurónových sietí.

Stratové funkcie sa líšia na základe povahy výstupov, ktoré má model predikovať. Príkladom stratovej funkcie pri regresných úlohách, kedy sa predikuje odhad číselných výstupov na základe nezávislých premenných, je Mean Squared Error (MSE). Danú funkciu definujeme ako [9]:

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (2.7)$$

Kde:  $N$  predstavuje počet skúmaných vzoriek,  $Y_i$  predstavuje cieľovú predikciu  $i$  – tej vzorky a  $\hat{Y}_i$  je predikciu modelu pre  $i$  – tu vzorku.

Pri viactriednych klasifikačných úlohách sa využíva stratová funkcia *Softmax Loss*, tiež nazývaná *Cross-Entropy Loss* (presnejšie *Categorical Cross-Entropy Loss*)[65]:

$$L_{Softmax} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{i,j} \log(p_{i,j}) \quad (2.8)$$

Kde:  $N$  predstavuje počet skúmaných vzoriek,  $C$  reprezentuje celkový počet tried,  $y_{i,j}$  predstavuje cieľové označenie  $i$  – tej vzorky pre  $j$  – tu triedu a  $p_{i,j}$  je predikovaná pravdepodobnosť  $i$  – tej vzorky pre  $j$  – tu triedu.

Pri úlohách binárnej klasifikácie, kde vzorky majú iba dve možnosti výsledku, sa používa *Binary Cross-Entropy Loss* [65]:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (2.9)$$

Kde:  $N$  predstavuje počet skúmaných vzoriek,  $y_i$  je cieľové označenie  $i$  – tej vzorky a  $p_i$  je predikované označenie  $i$  – tej vzorky.

Rýchlosť učenia (*learning rate*) určuje veľkosť kroku, ktorý optimalizačný algoritmus vykoná počas gradientálneho zostupu na aktualizáciu parametrov modelu. Ak je rýchlosť učenia príliš vysoká, model nemusí konvergovať, ale oscilovať okolo minima stratovej funkcie. Naopak, ak je rýchlosť učenia príliš nízka, tréningový proces môže byť pomalý, alebo sa model môže zaseknúť v lokálnych minimách.

Adam (*Adaptive Moment Estimation optimizer*) [39] je optimalizačný algoritmus, ktorý je známy vďaka svojej schopnosti rýchlej konvergencie k optimálnym hodnotám a tendencie robustnosti voči rôznym nastaveniam hyperparametrov<sup>4</sup>. Kombinuje vlastnosti stochastického gradientálneho zostupu (SGD) s adaptívnou rýchlosťou učenia pre každý parameter. Adam je taktiež schopný prispôbiť sa rôznym charakteristikám dát a ich distribúcii, čo vedie k stabilnému a efektívnemu učeniu aj pri väčších a komplexnejších dátových sadách.

Počet vzoriek počas jednej iterácie tréningovania, po ktorej sa aktualizujú váhy modelu, predstavuje pojem *Batch Size*. Epocha predstavuje jednu úplnú iteráciu cez celú tréningovú

<sup>4</sup>Hyperparametre sú externé konfiguračné nastavenia modelu definované pred tréningovým procesom, ktoré ovplyvňujú správanie a výkon modelu počas tréningu. Príklady hyperparametrov zahŕňajú počet epoch, *batch size*, *learning rate* a pod.

sadu dát a celkový počet epoch je teda počet iterácií, koľkokrát bola celá tréningová sada propagovaná cez neurónovú sieť.

## Hlboké učenie

Hlboké učenie predstavuje špecifickú oblasť strojového učenia, ktorá využíva neurónové siete na modelovanie a riešenie komplexných úloh. Jeho významným znakom je používanie hlbokých neurónových sietí, ktoré obsahujú viacero skrytých vrstiev. Tieto siete sú schopné naučiť sa reprezentácie dát na rôznych úrovniach abstrakcie, čo im umožňuje efektívne zvládať úlohy rôznych zložitostí.

Konkrétne varianty hlbokých neurónových sietí, ako sú konvolučné neurónové siete (CNN) a rekurentné neurónové siete (RNN), sa osvedčili v rôznych oblastiach vrátane počítačového videnia, spracovania prirodzeného jazyka a rozpoznávania reči. Ich schopnosť extrahovať príznaky a vzorov z obrazových, textových a zvukových dát ich robí vhodnými pre širokú škálu úloh v oblasti umelej inteligencie.

V prípade zachytávania priestorových vzorov z obrazových dát sú využívané konvolučné neurónové siete. Konvolúcia je matematická operácia zachytávajúca presah medzi dvomi funkciami. V kontexte konvolučných neurónových sietí konvolúcia zahŕňa „posúvanie“ konvolučného filtra cez vstupné dáta, pomocou čoho sa vypočíta bodový súčin medzi lokálnymi oblasťami vstupných dát [15].

Konvolučné siete sú ideálnym nástrojom pre klasifikáciu, segmentáciu obrázkov a detekciu objektov v nich [40]. Ich schopnosť zachytiť hierarchické vzory a štruktúry v obrázkových dátach im umožňuje dosahovať vysokú úspešnosť aj v komplexných úlohách počítačového videnia.

## ResNet

Počas tréningovania hlbokých neurónových sietí, najmä pri architektúrach s vysokým počtom skrytých vrstiev sa vyskytuje tzv. problém miznúceho gradientu (*vanishing gradient*). Tento problém nastáva počas tréningu pri spätnej propagácii gradientov z koncových vrstiev do počiatočných, keď gradienty nadobudnú primálne hodnoty (blížiac sa k 0). Výsledkom tohto problému je, že váhy počiatočných vrstiev prijímajú počas optimalizácie zanedbateľné aktualizácie, čo bráni v procese učenia.

ResNet (Residual Network) [29] je architektúra hlbokého učenia navrhnutá ako riešenie tohto problému. Táto architektúra predstavila jednoduchú, no inovatívnu ideu pridania skratkových pripojení (*skip connection*) zo vstupu na výstup aktivačnej funkcie.

Skratkové pripojenie, taktiež nazývané ako mapovanie identity, je priame prepojenie medzi vrstvami v neurónovej sieti. Tieto pripojenia obchádzajú jednu alebo viacero vrstiev pridaním vstupu predchádzajúcej vrstvy k výstupu neskoršej vrstvy, čím tvoria reziduálny blok. Fungovanie bloku so skratkovým pripojením môžeme jednoducho vyjadriť ako [29]:

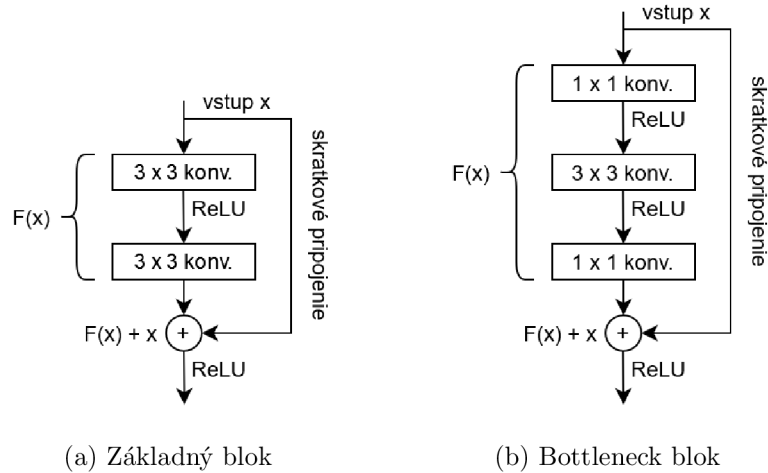
$$y = F(x) + x \tag{2.10}$$

Kde:  $y$  je výstupom reziduálneho bloku,  $F(x)$  predstavuje transformáciu aplikovanú daným blokom a  $x$  je vstupom bloku.

Tieto bloky umožňujú lepšiu spätnú propagáciu gradientov cez hlboké siete, čo zvyšuje ich schopnosť naučiť sa zložité vzťahy. Uľahčujú optimalizáciu sietí a zároveň fungujú ako forma regularizácie (udržiavajú informácie z pôvodného vstupu, čím zabezpečujú zachova-

nie podstatných informácií v priebehu tréovania), čo vedie k výrazne lepším výsledkom pri riešení komplexných úloh učenia v porovnaní s tradičnými hlbokými architektúrami.

ResNet pozostáva z dvoch hlavných reziduálnych blokov: základného bloku 2.6a, používaného v ResNet-18 a ResNet-34, a bottleneck bloku 2.6b, ktorý je súčasťou hlbších variácií, ako napríklad ResNet-50 a ResNet-152.



Obr. 2.6: Reziduálne bloky architektúry ResNet.

Základný blok pozostáva z dvoch konvolučných vrstiev s menším rozmerom konvolučného jadra (typicky  $3 \times 3$ ) a rovnakým počtom vstupných a výstupných kanálov. Za každou z týchto vrstiev nasleduje vrstva vykonávajúca batch normalization a aktivačná funkcia ReLU. Skratkové pripojenie priamo pridáva vstup na výstup druhej konvolučnej vrstvy.

*Bottleneck* blok Je navrhnutý k redukcii nákladnosti výpočtov a pozostáva z troch konvolučných vrstiev. Prvá  $1 \times 1$  konvolučná vrstva redukuje dimenzionalitu dát. Za ňou nasleduje  $3 \times 3$  konvolučná vrstva, a posledná  $1 \times 1$  konvolučná vrstva slúži k navráteniu redukovanej dimenzionality. Podobne ako pri základných blokoch, za každou konvolučnou vrstvou nasleduje *batch normalization* a ReLU aktivačná funkcia. Skratkové pripojenie pridáva vstup na výstup poslednej konvolučnej vrstvy.

Autori tohto modelu pôvodne vytvorili sieť s 34 vrstvami, inšpirovanú architektúrou VGG[63]. Následne pridali skratkové pripojenia na vytvorenie hlbšej siete, čo viedlo k zlepšeniu výkonu nielen vďaka navýšeniu parametrov, ale aj prehĺbením architektúry a zavedením skratkových pripojení.

Kľúčovým úspechom architektúry ResNet je jeho schopnosť prekonať problém degradácie bežne pozorovaný v hlbokých neurónových sieťach. Skratkové pripojenia umožňujú sieť konvergovať rýchlejšie ako pri obyčajných sieťach. Pôvodné architektúry tohto modelu s 18 a 50 vrstvami sú zobrazené v tabuľke 2.1.

Tabuľka 2.1: Architektúry ResNet-18 a ResNet-50

Názov Vrstvy	Rozmery Výstupu	RN-18	RN-50
conv1	112x112	$7 \times 7, 64, \text{stride } 2$	
		$3 \times 3 \text{ max pool, stride } 2$	
conv2_x	56x56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28x28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4_x	14x14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
conv5_x	7x7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1x1	avg pool, 1000-d fc, softmax	

## Face embedding

*Face embedding* [57] je technika využívaná v systémoch rozpoznávania tváre slúžiaca na reprezentáciu obrazov tváre ako číselných vektorov v priestore s vysokým počtom dimenzií (zvyčajne 128). Cieľom tejto techniky je zachytiť unikátne rysy a charakteristiky tváre takým spôsobom, ktorý umožňuje efektívne porovnávanie a rozpoznávanie.

*Face embedding* je často realizovaný pomocou modelov hlbokého učenia, najmä konvolučných neurónových sietí (CNN), ktoré sú trénované na rozsiahlych datasetoch obsahujúcich označené obrázky tváří. Tieto modely sú trénované k extrakcii relevantných črt zo snímok tváří, ako sú usporiadanie orientačných bodov tváre, textúry a ďalšie charakteristiky. Obsahujú špecifickú embedding vrstvu, ktorá generuje číselnú reprezentáciu pre každý vstupný obrázok tváre. Táto vrstva je trénovaná na mapovanie tváří do vysokorozmerného priestoru, kde sú podobné tváre mapované na blízke body a odlišné tváre na vzdialené body. Výstupom je číselný vektor, ktorý zachytáva základné črty tváre.

Počas úloh rozpoznávania tváří sa podobnosť medzi *embedded* tvármi počíta pomocou metrick, ako je napríklad euklidovská vzdialenosť alebo kosínusová podobnosť, čo umožňuje efektívne porovnávanie a zvyšuje presnosť rozpoznávania.

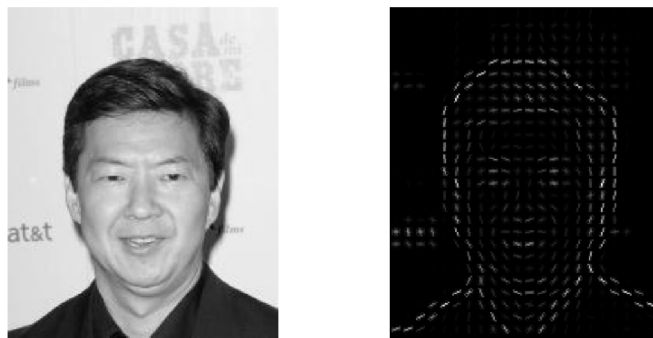
## 2.3 Algoritmy pre detekciu tváre

V oblasti detekcie tváří sa využívajú rôzne algoritmy na identifikáciu a lokalizáciu tváří v obrázkoch. Tieto algoritmy majú kľúčový význam v oblasti počítačového videnia a sú dôležitou súčasťou aplikácií ako je rozpoznávanie tváří alebo spracovanie obrazu.

### Histogram orientovaných gradientov

Aj keď *Histogram of oriented gradients* (HOG) [16] sám o sebe nie je algoritmom strojového učenia, používa sa pri predspracovaní obrazových dát v spojení s algoritmi strojového učenia pri účeloch detekcie objektov. HOG je deskriptor poskytujúci reprezentáciu štruktúry obrazu, ktorý zdôrazňuje vzory lokálnych gradientov.

Základnou myšlienkou HOG je, že vzhľad a tvar objektu v obrázku možno popísať tým, ako sa mení intenzita farieb alebo smer hrán. Proces začína rozdelením obrázka na malé oblasti, nazývané bunky. Pre každú bunku sa zostaví histogram smerov gradientu, čo reprezentuje, ako sa mení intenzita farieb v danej oblasti. Deskriptor potom vzniká zretazením týchto histogramov, čím vytvára kompaktný popis vzhľadu objektu. Na zvýšenie presnosti analýzy sa lokálne histogramy normalizujú kontrastom, ktorý sa vypočíta na základe intenzity naprieč väčšej oblasti. Táto normalizácia sa aplikuje na všetky bunky, čo vedie k lepšej schopnosti odolávať zmenám v osvetlení a zároveň zachováva citlivosť na dôležité detaily objektu. Príklad HOG je ilustrovaný na obrázku 2.7.



Obr. 2.7: Príklad HOG z obrázka

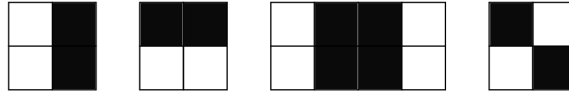
### Viola-Jones

Viola-Jones[67] je algoritmus detekcie tváre predstavený v roku 2001. Využíva Haar príznaky, štruktúru integrálneho obrazu, AdaBoost algoritmus a vytvára kaskády klasifikátorov k efektívnej a presnej selekcii črt tváre.

Haar príznak je jednoduchý pravouhlý obdĺžnikový filter používaný na zachytenie určitých vizuálnych vzorov na obrázku (najmä tváří v obrazových dátach). Dané filtre sa aplikujú na určitú časť obrazu a merajú rozdiel v intenzitách svetlých a tmavých oblastí.

Každý Haar príznak je definovaný svojou polohou, veľkosťou a váhou. Príznaky môžu zahŕňať jednoduché vzory, ako sú hrany, línie a zmeny kontrastov (obr. 2.8). Ich kombinácia je následne použitá pri trénovaní kaskádového klasifikátora, čím sa znižuje miera chýbajúcich alebo falošných detekcií [71]. Tento prístup je významným prvkom algoritmu detekcie tváří Viola-Jones.





Obr. 2.8: Príklady prvých Haar Príznakov pri Viola-Jones

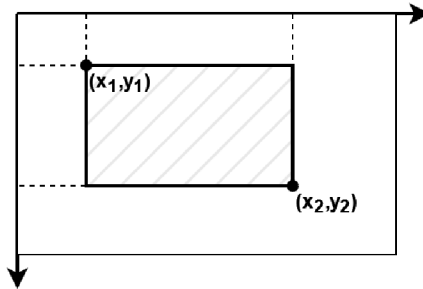
Viola-Jones algoritmus je navrhnutý tak, aby pracoval s čiernobielymi obrázkami, čo zjednodušuje výpočet Haar príznakov, ktoré sú kľúčovým komponentom algoritmu.

Haar príznaky zachytávajú určité vizuálne vzory na obrázku a predstavujú variácie v intenzite pixelov. Na urýchlenie výpočtu týchto príznakov sa používa integrálny obraz. Integrálny obraz je dátová štruktúra používaná na efektívny výpočet súčtu hodnôt pixelov v ľubovoľnej pravouhlej oblasti obrazu. Táto štruktúra môže byť definovaná ako dvojdimenzionálna vyhľadávacia tabuľka, v ktorej výpočet sumy intenzity na súradniciach  $(x, y)$  môžeme definovať ako:

$$s(x, y) = i(x, y) + s(x - 1, y) + s(x, y - 1) - s(x - 1, y - 1) \quad (2.11)$$

Kde:  $s(i,j)$  je suma intenzít pixelov od súradníc  $(0,0)$  po  $(i,j)$  a  $i(x,y)$  je intenzita pixelu na súradniciach  $(x,y)$

Na základe hodnôt vo vyhľadávacej tabuľke je možné vypočítať požadovaný súčet hodnôt za použitia iba štyroch hodnôt z tabuľky. V prípade obdĺžnika definovaného ľavým horným rohom  $(x_1, y_1)$  a pravým dolným rohom  $(x_2, y_2)$  môže byť suma hodnôt pixelov vnútri daného obdĺžnika vypočítaná nasledovne (obr. 2.9):



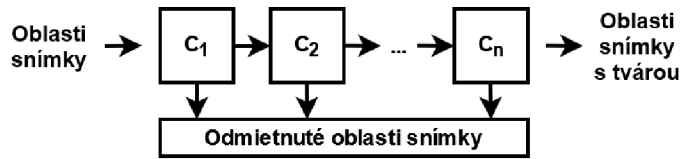
Obr. 2.9: Hľadanie sumy obdĺžnikovej oblasti

$$\sum_{rect} = s(x_2, y_2) - s(x_1 - 1, y_2) - s(x_2, y_1 - 1) + s(x_1 - 1, y_1 - 1) \quad (2.12)$$

Kde:  $\sum_{rect}$  je suma hodnôt pixelov vnútri obdĺžnika,  $s(i,j)$  je suma intenzít pixelov od súradníc  $(0,0)$  po  $(i,j)$ .

Jedná sa o výpočet v konštantnom čase, ktorý je značne rýchlejší ako iterácia cez všetky pixely v obdĺžniku a sčítavanie ich hodnôt jednotlivo. K výberu malej skupiny najdiskriminatívnejších Haar príznakov sa využíva Adaboost, kde sú tréňované slabé klasifikátory na konkrétne vlastnosti.

Vybrané vlastnosti a slabé klasifikátory sú usporiadané do kaskádovej štruktúry pozostávajúcej z viacerých úrovní, z ktorých každá obsahuje sadu klasifikátorov. Táto kaskáda umožňuje rýchle odmietnutie oblastí bez tváre na obrázku 2.10.



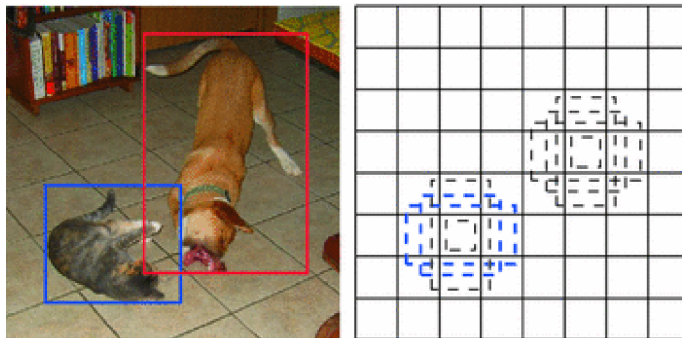
Obr. 2.10: Kaskáda slabých klasifikátorov algoritmu Viola-Jones

Viola-Jones algoritmus je široko používaný v počítačovom videní pre detekciu objektov. Techniky implementované v tomto algoritme ho robia vysoko efektívnym aj pri sekvencii obrázkov generovaných v reálnom čase zo štandardnej webovej kamery [69]. Daný algoritmus položil základ pre ďalší vývoj v oblasti detekcie objektov a zostáva nosným pilierom v tejto oblasti.

### Single Shot Multibox Detection (SSD)

SSD [42] je regresný algoritmus na detekciu objektov, ktorý využíva konvolučné neurónové siete na priamu predikciu polohy objektu a priradenie jeho triedy. Ide o metódu jedностupňovej detekcie, čo znamená, že vykonáva lokalizáciu aj klasifikáciu jedným prechodom siete, narozdiel od dvojstupňových detektorov, ako je Fast R-CNN [22], ktoré najprv detekujú oblasti záujmu, a až potom tieto oblasti klasifikujú.

Pojem *multibox* v SSD sa vzťahuje na sady predvolených ohraničujúcich rámcov, na detekciu objektov v rôznych mierkach a pomeroch strán. Tieto ohraničujúce rámčeky sú preddefinované tvary, ktoré pokrývajú rôzne oblasti vstupného obrazu. SSD predpovedá skóre tried pre každý ohraničujúci rámček. K tomu, aby poskytol výsledky detekcie, využíva algoritmus NMS (*Non-Maximum Suppression*) [6], na odstránenie duplicitných detekcií a výber najspoľahlivejších detekcií pre každú triedu objektov (obr. 2.11).



Obr. 2.11: Detekcia objektov pomocou SSD algoritmu [42].

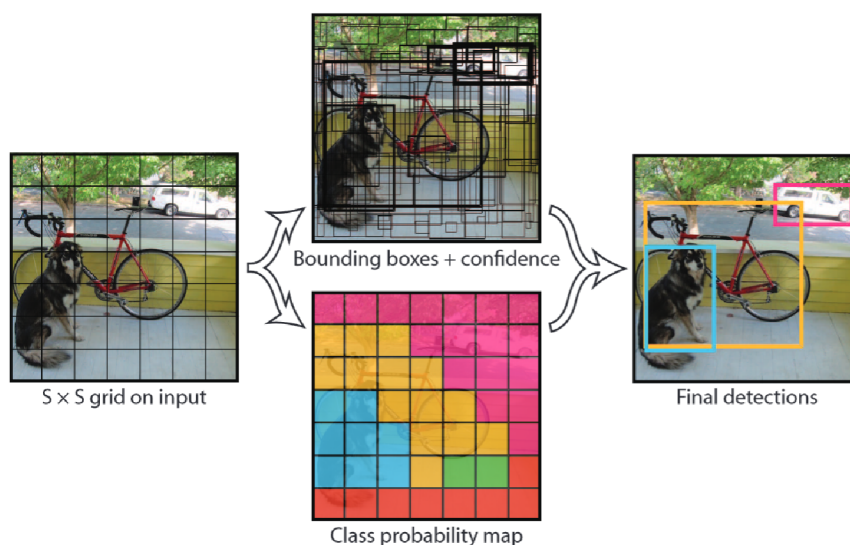
Počas trénovania modelu sa využíva kombinácia stratových funkcií pre lokalizáciu objektu (*Smooth L1 Loss* [22]) a klasifikáciu objektu (*Softmax Loss* alebo *Cross-Entropy Loss*).

Je dôležité poznamenať, že SSD je všeobecne použiteľný prístup detekcie objektov efektívny pri úlohách detekcie v reálnom čase. V dnešných systémoch detekcie tvárí sa často využívajú špecializovanejšie modely, napríklad dodatočným použitím algoritmov sledovania objektov alebo zmenou pôvodnej architektúry[33].

## You Only Look Once (YOLO)

YOLO [54] je algoritmus detekcie objektov predstavený v roku 2015. Základnou myšlienkou YOLO je rozdeliť vstupný obrázok do mriežky a pre každú bunku mriežky predpovedať hranice ohraničenia, skóre spoľahlivosti a pravdepodobnosti tried.

Každá bunka v mriežke predstavuje oblasť obrázka, ktorá bude použitá pri detekcii objektu. Pre každú bunku v mriežke sa predpovedá pevne stanovený počet ohraničujúcich rámečkov a súvisiace pravdepodobnosti tried. Predikcia každého ohraničujúceho rámečka pozostáva zo súradníc, ktoré definujú pozíciu, výšku a šírku rámečka vzhľadom na bunky mriežky, miera istoty (*confidence*), že rámeček obsahuje objekt a presnosť predikcie umiestnenia a triedy objektu (obr. 2.12).



Obr. 2.12: Detekcia objektov pomocou YOLO algoritmu [54].

Počas tréovania modelu sa využíva kombinácia stratových funkcií pre lokalizáciu objektu (*Smooth L1 Loss*), *confidence*, že ohraničujúci rámeček obsahuje objekt (*Binary Cross-Entropy Loss*) a klasifikáciu objektu (*Cross-Entropy Loss*).

YOLO vyniká svojou rýchlosťou a schopnosťou detekcie viacerých objektov v jedinom prechode, čo ho robí vhodným pre aplikácie v reálnom čase. Je však dôležité poznamenať, že v niektorých scenároch môže mať nižšiu presnosť v porovnaní s trochu pomalšími, ale presnejšími detektormi (ako napríklad SSD). Treba tiež zdôrazniť, že existujú rôzne verzie YOLO, pričom každá môže obsahovať vylepšenia alebo úpravy na zvýšenie výkonu.

## RetinaFace

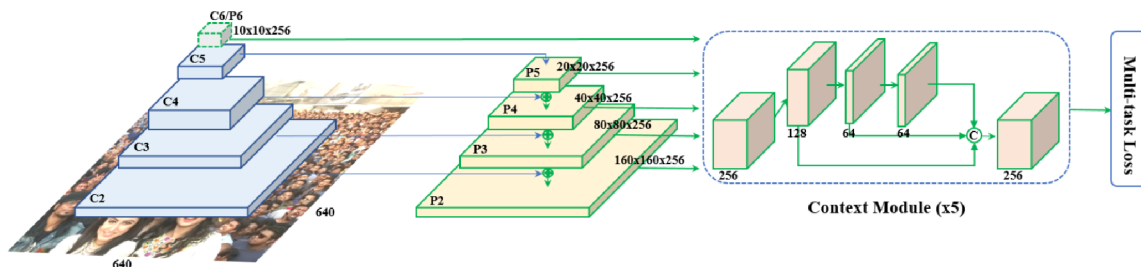
RetinaFace[11] je algoritmus hlbokého učenia navrhnutý na presnú detekciu tváří vrátane ohraničenia tváre a identifikácie orientačných bodov tváre, ako sú stredy očí, hrany pier a stred nosa.

Využíva tzv. *Feature Pyramid Network* (FPN) architektúru, ktorá bola navrhnutá tak, aby efektívne extrahovala viacškálové *features* z obrázkov. Výkon detekcie je zvýšený začlenením *feature* mapovaní z rôznych vrstiev siete, čo umožňuje detekciu objektov rôznych veľkostí. Na spojenie *features* z obrázkov FPN prepojí tieto vrstvy zhora nadol, čím sa vytvára viacúrovňová *feature* pyramída.

V RetinaFace sú úrovně pyramid (P2 až P5) odvodené od výstupu reziduálnych stupňov architektúry ResNet-152 (C2 až C5). V podstate sa informácie z vrstiev s vyšším rozlíšením kombinujú so sémanticky bohatými *features* z hlbších vrstiev, aby sa vytvorili reprezentácie funkcií vo viacerých mierkach. Okrem toho sa úroveň P6 vypočítava samostatne pomocou konvolučnej operácie 3x3 s krokom 2 aplikovaným na výstup konečného stupňa ResNet (C5).

Na každej z piatich úrovní pyramidy je implementovaný tzv. kontextový modul. Tieto moduly používajú techniku nazývanú deformovateľná konvolúcia, aby sieť lepšie pochopila komplikované a meniace sa pozadie objektov, ktoré sa snaží odhaliť. Kontextové moduly zaisťujú, že sa sieť nezameriava len na samotné objekty, ale zohľadňuje aj ich okolie, čo vedie k presnejšej detekcii.

*Multitask loss* v Retinaface je kombináciou stratových funkcií používaných počas tréningu pre optimalizáciu viacerých úloh súčasne. Tieto úlohy zahŕňajú predpovedanie súradníc ohraničujúceho rámčeka, súradnice orientačných bodov tváre a skóre spoľahlivosti detekcie tváří.



Obr. 2.13: Architektúra RetinaFace [11].

*Multitask loss* pozostáva z komponentov pre regresiu ohraničujúceho rámčeka (*Smooth L1 Loss*), regresiu kľúčových bodov tváre (*Smooth L1 Loss*) a predikciu *confidence* skóre (*Binary Cross-Entropy Loss*).

V porovnaní s predchádzajúcimi algoritmi YOLO a SSD, RetinaFace sa vyznačuje komplexnejšou architektúrou špecializovanou výhradne na detekciu tváre. Táto špecializácia vedie k vyššej presnosti, na úkor rýchlosti detekcie. Napriek tomu môže byť algoritmus RetinaFace vhodný aj pre aplikácie vyžadujúce detekciu v reálnom čase pri dostatočných výpočtových prostriedkoch.

## 2.4 Algoritmy pre rozpoznávanie tváre

Prvý vypracovaný výskum zaoberajúci sa automatizovaným rozpoznávaním tváre bol uskutočnený už v 60. rokoch 20. storočia. Išlo o výskum využívajúci hlavne geometriu tváre [5]. Vstupom bola poloautomatická metóda, pri ktorej mali operátori pomocou RAND tabletu zadať 20 mier, ako napríklad stred zreničiek, kútiky očí a šírka úst. Touto metódou sa dalo spracovať okolo 40 snímok za hodinu. Vypočítané vzdialenosti z fotografie boli spojené so záznamom mena osoby. Vo fáze rozpoznávania sa porovnávali vzdialenosti na fotografiách so vzdialenosťami uložených v databáze, a navrátili sa najbližšie záznamy. V praxi táto metóda nebola uplatnená.

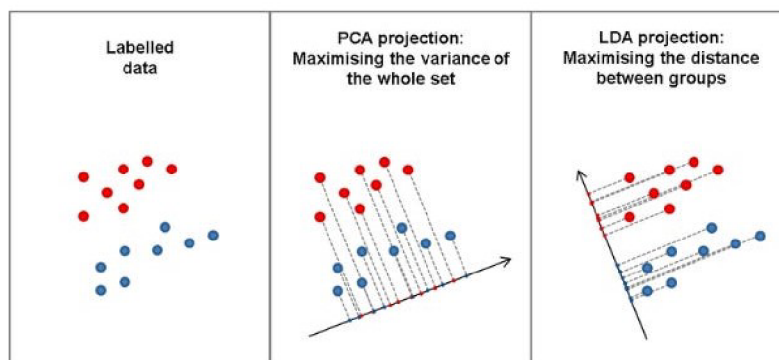
## Principal Component Analysis (PCA)

Analýza hlavných komponentov [4] (PCA) je nekontrolovaná štatistická metóda redukcie dimenzionality dát, ktorá bola v roku 1991 rozšírená na rozpoznávanie tvárí. PCA sa zvyčajne používa na predbežné spracovanie údajov pred inými analýzami. Tento prístup znižuje rozmery dát pomocou kompresie a odhaľuje najefektívnejšiu nízko-dimenzionálnu štruktúru črt tváre (*features*). Táto redukcia odstraňuje redundantné informácie a šum, zatiaľ čo zachováva základné charakteristiky údajov, výrazne znižuje rozmery a urýchľuje spracovanie dát.

Zatiaľ čo existujú pokročilejšie techniky rozpoznávania tvárí (napr. prístupy hlbokého učenia), PCA slúži ako základná technika redukcie dimenzionality v tejto doméne.

## Linear Discriminant Analysis (LDA)

Lineárna diskriminančná analýza [4] (LDA) je kontrolovaná technika separácie dát. Cieľom LDA je nájsť smer, ktorý vhodne rozdelí rôzne triedy dát po ich projekcii. LDA úzko súvisí s PCA v tom, že obe hľadajú lineárne kombinácie premenných, ktoré najlepšie popisujú dáta. Na rozdiel od PCA, ktorej cieľom je maximalizovať zníženie rozptylu údajov, LDA sa pokúša minimalizovať rozptyl v rámci rovnakej kategórie údajov pri projekcii a súčasne maximalizovať rozptyl medzi odlišnými kategóriami (obr. 2.14).



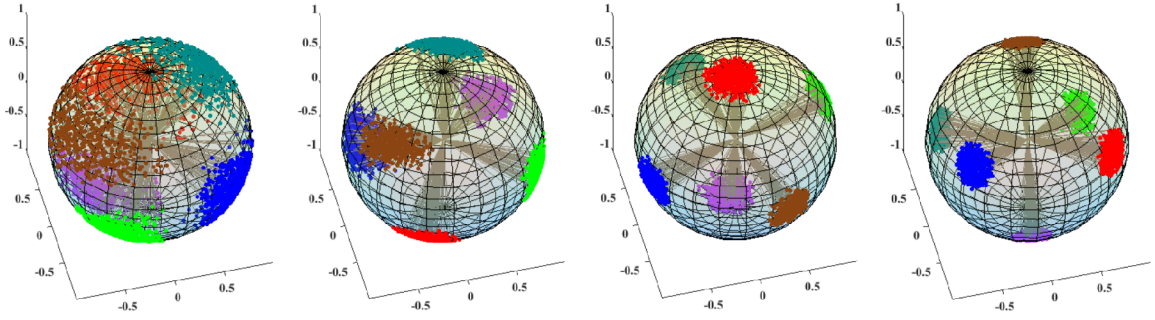
Obr. 2.14: Porovnanie PCA a LDA projekcií [48]

Kombinácia PCA a LDA (PCA-LDA) poskytuje súbor rozlišovacích funkcií, ktoré sú vhodné na rozpoznávanie tváre. Taktiež umožňujú algoritmom rýchlejšie a efektívnejšie pracovať, pretože obidve znižujú časovú zložitosť spracovania dát. Trénovaním klasifikátora (napr. K-Nearest-Neighbour alebo SVM) pomocou funkcií extrahovaných procesom PCA-LDA, je možné daný klasifikátor použiť na rozpoznávanie tvárí.

## SphereFace

SphereFace[44] je algoritmus rozpoznávania tvárí, ktorý adresuje výzvy spojené s veľkými rozdielmi v pózach, výrazoch a osvetlení na snímkach. Jeho významným prínosom je schopnosť učiť sa rozlišovacie črty tvárí v tzv. *hypersphere manifold*, čo znamená, že vytvára kompaktné reprezentácie *embedded features* tváre na hyperkulovitej ploche (obr. 2.15).





Obr. 2.15: Rozloženie vzoriek tvárí pomocou algoritmu SphereFace [44].

Stratová funkcia metódy SphereFace je odvodená od *Angular Softmax Loss*, ktorá zavádza uhlovú rezervu, tzv. *angular margin* k štandardnej funkcii *Softmax Loss*, čím sa vytvárajú väčšie uhlové rozdiely medzi triedami vo *feature space*. *Angular Softmax Loss* môže byť vyjadrená ako [44]:

$$L_{A-Softmax} = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{\|x_i\| \psi(\theta_{y_i})}}{e^{\|x_i\| \psi(\theta_{y_i})} + \sum_{j=1, j \neq y_i}^N e^{\|x_i\| \eta(\theta_j)}} \quad (2.13)$$

Kde:  $N$  predstavuje počet vzoriek, na ktorých sa vykonáva výpočet stratovej funkcie,  $\|x_i\|$  predstavuje normalizovaný vektor *features*,  $\theta_{y_i}$  je uhol medzi *feature* vektorom a klasifikátorom,  $\psi(\theta_{y_i})$  je uhlová aktivačná funkcia pre cieľovú triedu a  $\eta(\theta_j)$ , kde  $j \neq y_i$  je uhlovou aktivačnou funkciou váhových vektorov triedy, ktorá nie je cieľová.

Zavedením parametru  $m$ , a prepísaním aktivačných funkcií  $\psi(\theta)$  a  $\eta(\theta)$  sa získa stratová funkcia *SphereFace Loss*, ktorá môže byť zjednodušene definovaná ako [43]:

$$L_{SphereFace} = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{s \cdot \cos(m \theta_{y_i})}}{e^{s \cdot \cos(m \theta_{y_i})} + \sum_{j=1, j \neq y_i}^N e^{s \cdot \cos(m \theta_j)}} \quad (2.14)$$

Kde  $s$  je polomerom hyperkulovitého priestoru a  $m$  je parametrom, ktorý kontroluje veľkosť *angular margin* medzi triedami vo *feature space*, čo umožňuje lepšiu separáciu tried vo vektorovom priestore. V pôvodnom SphereFace algoritme bol parameter  $m$  vyžadovaný ako celé číslo, avšak v praxi sa ukázalo, že to nie je nevyhnutné obmedzenie. Pre lepšiu stabilitu tréningu sa teda  $m$  optimalizuje spoločne so stratovou funkciou [43].

## ArcFace

ArcFace [12] je algoritmus v oblasti rozpoznávania tváre, vytvorený s cieľom zvýšiť schopnosť rozlišovania *embedded features* tváre, čím sa stávajú vhodnejšími pre verifikačné a identifikačné úlohy.

ArcFace algoritmus využíva stratovú funkciu *Additive Angular Margin Loss* (tiež nazývaná ako *ArcFace Loss*), ktorá sa opiera o stratovú funkciu *Angular Softmax Loss*. *ArcFace Loss* môže byť definovaná ako [12]:

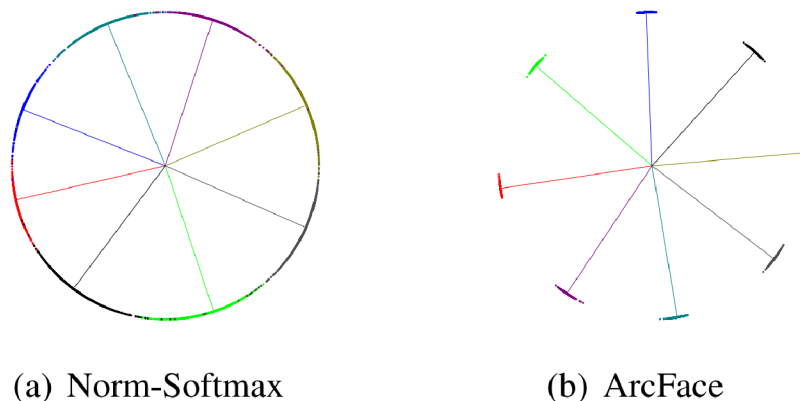
$$L_{ArcFace} = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{s \cos(\theta_{y_i} + m)}}{e^{s \cos(\theta_{y_i} + m)} + \sum_{j=1, j \neq y_i}^N e^{s \cos \theta_j}} \quad (2.15)$$

Kde parametre nesú rovnaký význam ako pri *SphereFace Loss*.

*ArcFace Loss* bola navrhnutá s cieľom zabezpečiť, že prvky rovnakej triedy budú v uhlovom priestore blízko seba, zatiaľ čo vlastnosti rôznych tried budú vhodne oddelené.

ArcFace presadzuje *angular margin* medzi triedami pridaním  $m$  ku kosínusu uhla, čo vytvára väčšie oddelenie medzi *embedded features* rozdielnych tried v uhlovom priestore.

Na ilustráciu tohto procesu autori trénovali svoj model pre rozlišovanie medzi 8 identitami pomocou stratovej funkcie Softmax a ArcFace (obr. 2.16). Bodky reprezentujú jednotlivé vzorky a čiary sú spojené so stredovým bodom každej identity. Vďaka normalizácii sú všetky črty tváre posunuté do oblúkového priestoru s pevným polomerom  $s$ .

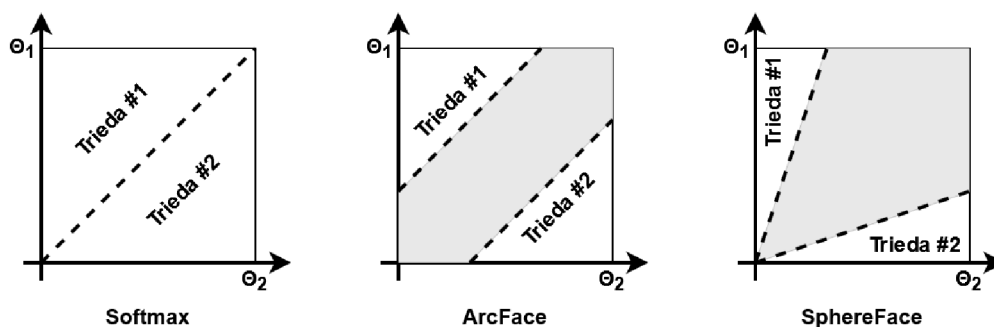


Obr. 2.16: Rozdiely pri použití stratovej funkcie Softmax a ArcFace [12].

Kľúčovým rozdielom medzi stratovými funkciami ArcFace a Sphereface spočíva v tom, ako dané funkcie narábajú s parametrom  $m$ , čiže s hodnotou používanou k separácii medzi jednotlivými triedami vo *feature space*.

SphereFace a ArcFace sú pokročilé metódy rozpoznávania tvárí, ktoré využívajú stratové funkcie založené na *angular margin*. Obe tieto techniky zlepšujú výkon v náročných situáciách rozpoznávania v porovnaní s predošlými metódami používajúce *Softmax Loss*.

Na obrázku 2.17 sú graficky znázornené rozdiely medzi jednotlivými metódami pri binárnej klasifikácii. Čiarkované línie reprezentujú rozhodovaciu hranicu, a sivé oblasti rozhodovacie rezervy.



Obr. 2.17: Rozdiely metód Softmax, ArcFace a SphereFace pri binárnej klasifikácii[12].





## 2.5 Etické a právne aspekty rozpoznávania tváre

Technológia rozpoznávania tváre predstavuje príležitosti aj výzvy pre základné práva. Aj keď môže zvýšiť bezpečnosť tým, že pomáha pri úlohách, ako je vyhľadávanie nezvestných ľudí a odhaľovanie podvodov, vyvoláva tiež obavy o súkromie a občianske slobody. Vnímanie rozpoznávania tváre sa v rámci populácie líši, pričom niektorí vyjadrujú nepohodlie alebo pocit rušivého používania. Prieskumy uskutočnené v rôznych kontextoch poukazujú na odlišné úrovne prijatia danej technológie v rámci populácie.

Pri implementácii kamerových systémov na účely rozpoznávania tváre vzniká viacero problematických aspektov, ktoré je dôležité zohľadniť. Technológia rozpoznávania tváre môže byť vnímaná ako invazívna, nakoľko umožňuje identifikáciu a sledovanie jednotlivcov bez ich vedomia alebo súhlasu. Zároveň existuje obava zo zneužitia biometrických údajov tváre na rôzne účely, vrátane krádeže identity na spáchanie podvodov alebo iných nezákonných aktivít. Vytváranie falošných identít na šírenie dezinformácií či manipuláciu verejným názorom predstavuje ďalší potenciálny problém. Rovnako tak sledovanie správania spotrebiteľov a cielený marketing zo strany veľkých firiem môžu vyvolať otázky týkajúce sa súkromia a etiky.

Pri navrhovaní a zavádzaní technológií rozpoznávania tváre sa musia rešpektovať základné práva vrátane ľudskej dôstojnosti. Dodržiavanie právnych predpisov EÚ a medzinárodného práva v oblasti ľudských práv je nevyhnutné na zabezpečenie toho, aby postupy rozpoznávania tváre boli v súlade s právami. To zahŕňa zváženie faktorov ako je presnosť, nevyhnutnosť a proporcionalita. Hoci technológie rozpoznávania tváre môžu zvýšiť efektivitu pri presadzovaní práva, obavy z chýb a falošných pozitív musia byť starostlivo vyvážené potrebou chrániť práva jednotlivcov.

Celkovo sa nasadenie technológií rozpoznávania tváre musí riadiť rešpektovaním základných práv a dodržiavaním právnych rámcov [20].

## Kapitola 3

# Hodnotenie kvality snímky

Schopnosť vyhodnotiť kvalitu snímok predstavuje kľúčový aspekt množstva aplikácií pracujúcich s obrazovými dátami. Integrácia algoritmov na hodnotenie kvality obrazu (*Image Quality Assessment* - IQA) [68] do systémov spracovania obrazu umožňuje presnejšiu interpretáciu údajov a zlepšuje výkon týchto systémov. IQA algoritmy majú tiež svoje uplatnenie v *Benchmark Image Processing* systémoch [14], kde je potrebné vyhodnotiť a vybrať najlepší systém spracovania obrazu pre konkrétnu úlohu.

V bezpečnostných aplikáciách umožňujú IQA algoritmy analýzu kvality obrazových dát, napríklad identifikáciou problémov ako rozmazanie alebo šum, ktoré môžu obmedziť efektivitu monitorovania. V multimediálnych aplikáciách môžu IQA algoritmy slúžiť na optimalizáciu kompresie obrazových dát tým, že automaticky určia úroveň kompresie, ktorá zachováva akceptovateľnú kvalitu obrazu, čím sa zefektívni proces uchovávanía a prenosu dát bez výrazného poklesu kvality.

### 3.1 Možnosti vyhodnocovania kvality

Na základe dostupnosti referenčných obrázkov, môžeme IQA metódy rozdeliť na: *full-reference* (FR), *reduced-reference* (RR) a *no-reference* (NR) [68].

Ďalším kľúčovým rozdelením IQA metrík je ich účel. Ak sú dané metódy zamerané na určenie kvality na základe vlastností ľudského zraku, hovoríme o *Human Visual System* (HVS) [70] orientovaných metrikách, ak je vyhodnocovanie kvality zamerané na špecifické úlohy rozpoznávania tváří, môžeme hovoriť o metrikách orientovaných na rozpoznávanie tváre.

#### 3.1.1 HVS orientované IQA metriky

Tradičné metriky hodnotia kvalitu snímok na základe vlastností ľudského vnímania. HVS orientované metriky sa často snažia modelovať rôzne aspekty ľudského zraku, ako sú kontrastové citlivosti, farebné vnímanie, ostrosť, vnímanie šumu a iné.

Každá z daných kategórií má svoje výhody a limitácie. Výber hodnotiacej metódy závisí na špecifických požiadavkách vykonávanej úlohy a dostupnosti pôvodných snímok. Treba avšak poznamenať, že cieľom tradičných IQA metód je vyhodnotiť kvalitu na základe vlastností ľudského vnímania, namiesto algoritmov rozpoznávania tváří, čo z nich robí suboptimálne riešenie pri hodnotení kvality pre rozpoznávacie účely.

## FR-IQA

Pri FR-IQA metódach sa porovnáva kvalita hodnoteného obrázku s kvalitou referenčného (pôvodného) obrázku. Tieto algoritmy zvyčajne pracujú s metrikami podobnosti medzi týmito dvomi obrázkami, ako sú napríklad *Structure Similarity Index* (SSIM) [31] a *Peak Signal to Noise Ratio* (PSNR) [31].

SSIM je percepčná metrika kvantifikujúca podobnosť medzi dvomi obrázkami na základe ich štrukturálnych informácií (porovnáva jas, kontrast a štruktúru obrázkov). Hodnota SSIM indexu sa vypočíta porovnaním priemerov, štandardnou odchýlkou a kovarianciou intenzít pixelov v malých oblastiach naprieč snímkami.

PSNR hodnotí kvalitu na základe množstva prítomného šumu alebo skreslenia medzi dvomi obrázkami. Meria pomer medzi maximálnou možnou hodnotou pixelu a silou šumu (kvantifikovaného pomocou MSE). Hodnota PSNR sa vyjadruje v decibeloch, pričom vyššie hodnoty naznačujú nižšie úrovne šumu alebo skreslenia oproti referenčnému obrázku. Táto metrika je používaná vďaka svojej jednoduchosti výpočtu, avšak interpretácia jej hodnôt by sa mala používať v spojení s inými metrikami (ako napríklad SSIM), pretože PSNR nemusí efektívne zachytiť degradácie kvality pri špecifických skresleniach, na ktoré je ľudský zrak citlivý.

Dané metriky sa používajú zväčša pri obrázkoch s odtieňmi šedej, avšak pri hodnotení farebných obrázkov je možné dané výpočty použiť separátne na každom farebnom kanáli. Vyhodnocovanie kvality danými spôsobmi má svoje využitie, iba ak je počas vyhodnocovania k dispozícii aj pôvodný obrázok, s ktorým vieme vykonať porovnanie (napríklad pri kompresii obrázkov).

## RR-IQA

Počas vyhodnocovania kvality RR-IQA metódami sa vyhodnotenie kvality obrazu posudzuje pomocou čiastočných informácií z pôvodného obrazu. Tieto algoritmy extrahujú relevantné vlastnosti alebo atribúty z hodnotených aj referenčných obrázkov a porovnávajú ich, aby kvantifikovali rozdiel v kvalite. Príklady týchto metód sú *Feature Similarity Index Measure* (FSIM) [73] a *Visual Saliency-induced Index* (VSI) [72].

FSIM vyhodnocuje podobnosť na základe vysokoúrovňových prvkov medzi dvoma obrázkami. Zameriava sa na extrakciu štrukturálnych vlastností ako sú hrany, textúry a gradienty a posudzuje ich podobnosť medzi obrázkami.

VSI hodnotí vizuálnu výraznosť obrázka na základe miery oblastí alebo objektov v obrázku, ktoré na seba priťahujú pozornosť. VSI kvantifikuje rozdiely tejto výraznosti medzi obrázkami, čím poskytuje mieru zachovania vizuálnej estetiky.

## NR-IQA

Keďže v mnohých aplikáciách počítačového videnia nie sú dostupné pôvodné referenčné obrázky, NR-IQA sa stala z najzložitejších úloh objektívneho hodnotenia kvality obrazu. Dané algoritmy predpovedajú kvalitu hodnotenej snímky iba na základe jej vlastností.

Príkladom NR-IQA je model *Naturalness Image Quality Evaluator* (NIQE), ktorý sa používa na určovanie kvality obrázkov na základe *Natural Scene Statistics* (NSS) [61]. Jeho cieľom je zachytenie charakteristík v obraze, ako je distribúcia hrán, textúrne vzory a korelácia farieb, odvodených z „prirodzených“ obrázkov. Na základe týchto charakteristík poskytuje hodnotenie kvality, ktoré sa zameriava na autentickosť obrázkov.

### 3.1.2 IQA metriky pri rozpoznávaní tváre

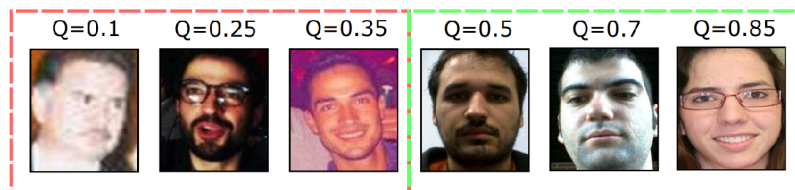
Metriky hodnotenia obrazu orientované na rozpoznávanie tváre sú špecificky navrhnuté tak, aby posudzovali ako dobre sú podstatné črty a charakteristiky tváre zachované alebo ovplyvnené vplyvom rôznych faktorov, ako napríklad: rozmazaním, šumom, kompresiou, zakrytím kľúčových črt, rôznymi svetelnými podmienkami alebo rotáciou tváre jednotlivca. Ich cieľom je poskytnúť prehľad o tom, ako kvalita obrazu ovplyvňuje úspešnosť algoritmov rozpoznávania tváre.

#### Reduced Reference Automatic Ranking (RRAR)

RRAR [74] je IQA algoritmus zlepšujúci výkon algoritmov rozpoznávania tváří. Pozostáva z dvoch častí: komponent kontroly kvality a komponent hodnotenia kvality. Pri kontrole kvality sa pomocou upravených SSIM a SVM algoritmov filtrujú nekvalitné obrázky tváří na základe osvetlenia a pózy tváre. Počas hodnotenia kvality sa využívajú výstupy upraveného SSIM skóre.

#### FaceQNet

FaceQNet[30] je RR-IQA algoritmus, ktorý využil už natrénovanú CNN na rozpoznávanie tváří. Daná sieť bola modifikovaná k porovnávaniu skóre medzi snímkami v rôznych podmienkach (vypočítané pomocou euklidovskej vzdialenosti) a snímkami vysokej kvality podľa štandardu ICAO [18]. Použitím tejto upravenej siete sa následne extrahovali kvalitatívne merania z rôznych datasetov (obr. 3.1). Tieto merania sa ukázali ako spoľahlivý odhad presnosti rozpoznávania.



Obr. 3.1: Hodnotenie kvality snímky tváre pomocou FaceQNet [30].

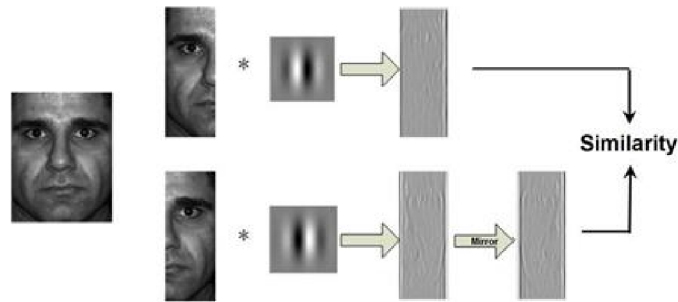
#### Gabor-Based Facial Symmetry and DCT-Based Sharpness

Sang a kolektív [55] navrhli určovanie kvality snímok pomocou Gáborových vlniek, diskretnej kosínusovej transformácie (DCT) a inverznej DCT (IDCT) na základe štandardov popísaných v sekcii 3.2.

Gáborove vlnky sú mocné nástroje používané na detekciu a extrakciu textúr a vzorov v obraze, vďaka ich schopnosti zachytiť lokálne vlastnosti s rôznymi orientáciami a škálami. Aplikáciou Gáborových filtrov na ľavú a pravú polovicu tváre a následným porovnaním výsledkov je možné identifikovať asymetriu v tvári (obr. 3.2).

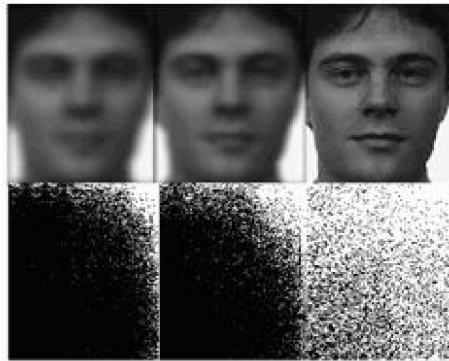
Tento proces umožňuje odhaľovať aj jemné rozdiely medzi polovicami tváre, ktoré môžu byť spôsobené rôznymi svetelnými podmienkami a pózami. Tieto informácie môžu poskytnúť dôležitý kontext pri hodnotení kvality obrazu a môžu byť využité pri spracovaní a analýze snímok tváre.

Transformáciou obrázku z priestorovej do frekvenčnej domény pomocou DCT sa informácie v obraze reprezentujú pomocou rôznych frekvenčných zložiek. Keďže vysokofrek-



Obr. 3.2: Vyhodnotenie asymetrie tváre a svetelných podmienok pomocou Gáborových vlniek [55].

venčné informácie sú spojené s detailmi a hrúbkou hrán, v prípade, že je obrázok rozostrený, dochádza k strate týchto informácií (obr. 3.3).



Obr. 3.3: Transformácia obrázkov do frekvenčnej domény pomocou DCT [55].

Po spätnej transformácii obrázku z frekvenčnej do priestorovej domény pomocou IDCT sa získa obnovený obrázok. Rozdielom vstupného a obnoveného obrázku sa určí ostrosť. Čím väčší je rozdiel, tým je pôvodný obrázok ostrejší.

## 3.2 Faktory ovplyvňujúce rozpoznávanie tváre

Rozpoznávanie tváre je komplexný proces ovplyvnený rôznymi faktormi, ktoré zahŕňajú nielen technické parametre systému, ale aj podmienky, za ktorých sú tváre zachytávané. Okrem toho je dôležité zohľadniť aj faktory ako svetelné podmienky, kvalita záberu, úroveň detailov, ktoré môžu ovplyvniť úspešnosť rozpoznávania tváre.

Na základe týchto podmienok môžeme snímanie rozdeliť do dvoch kategórií: snímanie v kontrolovanom prostredí a v reálnych podmienkach. Každá z týchto kategórií prináša svoje vlastné výzvy a úlohy, ktoré treba brať do úvahy pri návrhu a implementácii systémov rozpoznávania tváre. Zatiaľ čo kontrolované prostredie umožňuje lepšiu kontrolu nad podmienkami snímania, čo ho kvalifikuje pre biometrické účely, reálne podmienky často predstavujú väčšiu variabilitu v kvalite snímok, čo je potrebné zohľadniť pri vývoji spoľahlivých a efektívnych riešení rozpoznávania tváre [13].

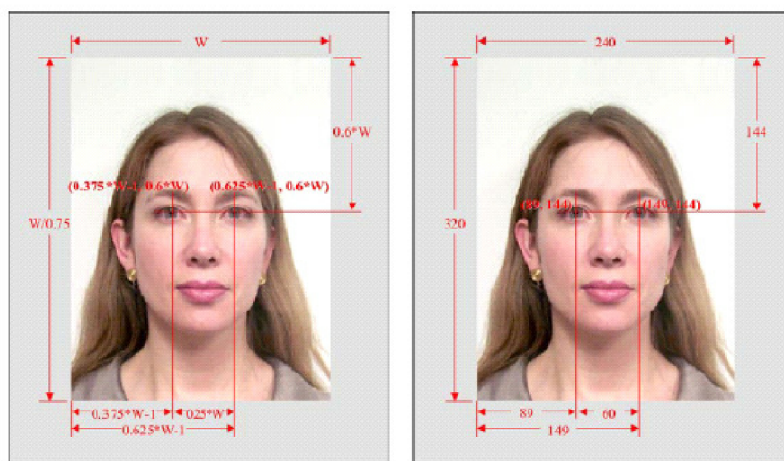
### 3.2.1 Kontrolované snímanie

Pri identifikácii jednotlivcov zo snímok tváre v kontrolovanom prostredí je možné minimalizovať vplyv mnohých faktorov, ktoré ovplyvňujú výsledky rozpoznávania tým, že umožňuje optimalizáciu svetelných podmienok a presné zameranie kamery na konkrétne oblasti tváre.

ISO/IEC 19794-5 [2] je medzinárodný štandard určený na definovanie formátu súborov pre biometrické dáta tváre. Poskytuje popis osvedčených postupov pri fotografovaní tváří a špecifikuje fotografické vlastnosti, obmedzenia scény a digitálne atribúty obrazov tváre používaných v biometrických aplikáciách, ako je rozpoznávanie tváre.

Táto norma umožňuje jednotnú štandardizáciu digitálnych obrazov tváre a ich výmenu medzi rôznymi platformami a aplikáciami. Zohráva kľúčovú úlohu pri zabezpečení kompatibility a spoľahlivosti v oblasti biometrických technológií, predovšetkým v kontexte identifikácie a overovania tváre.

Snímka schválená týmto štandardom (obr. 3.4) a požiadavky snímok spĺňajúcich tento štandard (tabuľka 3.1) s sú zobrazené nižšie.



Obr. 3.4: Štandardné parametre snímky tváre v súlade s ISO/IEC 19794-5:2005[34].

Tabuľka 3.1: Požiadavky snímky tváre v ISO/IEC 19794-5:2005[34].

Požiadavky	Atribúty	Obmedzenia
Scénické	Držanie tela	Kontrola odklonu od kamery
	Osvetlenie	Rovnomerné bez tieňa
	Pozadie	Jednoduché, svetlej farby
	Oči	Otvorené a jasne viditeľné
	Okuliare	Bez odrazov blesku a tmavého odtieňa
	Ústa	Zatvorené a jasne viditeľné
Fotografické	Pozícia hlavy	Umiestnená v strede
	Vzdialenosť kamery	Primeraná veľkosti hlavy
	Farba	Neutrálna a bez červených očí
	Expozícia	Primeraná jasnosti
Digitálne	Zaostrenie	Ostrá snímka bez rozostrení
	Rozlíšenie	Obmedzené šírkou hlavy

### 3.2.2 Snímanie v reálnych podmienkach

Problémy pri rozpoznávaní tváre môžu nastať v prípade fotografií pochádzajúcich z neideálnych podmienok. Takéto snímky môžu obsahovať rôzne degradácie, ktoré robia danú snímku nevhodnú pre účely rozpoznávania.

Na riešenie normatívnych požiadaviek ISO/IEC 19794-5 (popísanej vyššie), ktoré otvorili diskusiu o probléme objektívnej metriky pre kvalitu vnímanú človekom, štandard ISO/IEC 29794-5 [1] zaviedol metriky psFAR (Probabilistic False Acceptance Rate) a psFRR (Probabilistic False Rejection Rate). Tieto metriky predstavujú pravdepodobnosť falošného prijatia neoprávnenej osoby a pravdepodobnosť odmietnutia oprávnenej osoby v biometrickom autentifikačnom systéme. Ďalej špecifikuje metodológie na výpočet objektívneho a kvantitatívneho hodnotenia kvality pre snímky tváre. Navrhuje, aby sa faktory ovplyvňujúce kvalitu snímku tváre kategorizovali na statické a dynamické charakteristiky. Medzi statické charakteristiky patrí napríklad rozlíšenie snímok, textúra pozadia, existencia okuliarov v rámci snímky alebo dokonca špecifické biologické vlastnosti jedinca (napr. jazvy alebo znamienka). Pri dynamických charakteristikách ide o dočasné javy, ako rozostrenie kamery, variácie v osvetlení a póza alebo výraz tváre jedinca. Štandard zavádza prístupy pri určovaní aspektov ako je symetria tváre, rozlíšenie, intenzita osvetlenia, jas, kontrast, expozícia, ostrosť a pod.

Karahan a kolektív [36] sa zaoberali analýzou vplyvu rôznych degradačných faktorov na výkon hlbokých neurónových sietí, ktoré sa používajú na rozpoznávanie tváre. V štúdiu bola použitá kosínusová vzdialenosť na meranie podobnosti medzi obrázkami.

Obrázky boli degradované pomocou rôznych techník, vrátane rozmazania gaussovým a mediánovým filtrom, pridania gaussového a *Salt and Pepper* šumu, zmeny kontrastu vynásobením hodnôt farebných kanálov, skreslenia farieb zvýšením intenzity jednotlivých farebných kanálov a zakrytia rôznych častí tváre tmavými elipsovými alebo obdĺžnikovými tvarmi.

Výsledky tejto štúdie naznačujú, že testované modely dosahovali najslabšie výsledky pri rozmazaných obrázkoch, obrázkoch s vysokým šumom a obrázkoch, kde boli prekryté kľúčové oblasti tváre (oči, nos a ústa). Naopak, modely boli odolné voči zmenám kontrastu a skresleniu farieb.

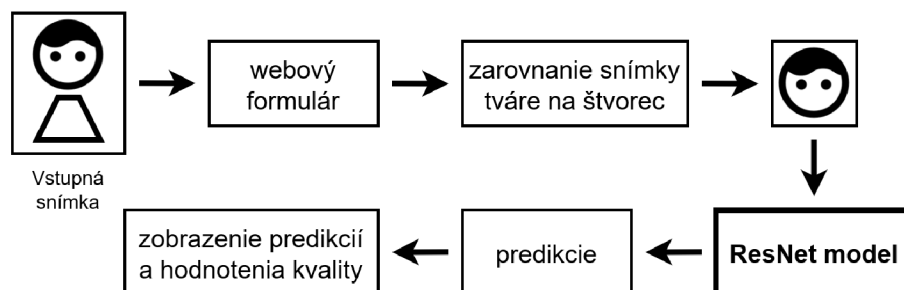
## Kapitola 4

# Návrh a implementácia

Cieľom aplikácie je spracovať vstupnú snímku od užívateľa a vyhodnotiť kvalitu danej snímky pomocou modelu neurónovej siete s architektúrou ResNet. Prvým krokom pri návrhu aplikácie je definícia jednotlivých krokov potrebných k dosiahnutiu požadovaných výsledkov.

Aplikácia by mala byť schopná spracovať akúkoľvek snímku v relevantnom formáte (JPEG, PNG alebo WebP) prostredníctvom vstupného webového formulára. Na danej snímke je potrebné vykonať detekciu tváre a oblasť, v ktorej sa tvár nachádza, zarovnať na štvorec. Veľkosť tejto štvorcovej oblasti je nevyhnutné upraviť tak, aby odpovedala vstupu, ktorý vyžaduje natrénovaný model. Štvorcová oblasť je následne odoslaná ako vstup do modelu, ktorý na základe jej vlastností vykoná predikciu, podľa ktorej sa vypočíta kvalita vstupnej snímky tváre. Výsledné vyhodnotenie snímky je následne zobrazené užívateľovi v podobe skóre kvality a tabuľky predikcií.

Diagram znázorňujúci tento návrh aplikácie je uvedený na obrázku 4.1.



Obr. 4.1: Diagram navrhnutej aplikácie.

### 4.1 Príprava datasetu

Pre optimálne fungovanie modelu je dôležité mať dostatočne veľký dataset s tvármi rôznej kvality. Tento cieľ môžeme dosiahnuť pridaním vhodných fotografií z existujúcich dátových sád a ich prípadnou úpravou. Takýto dataset, označovaný ako FIQVD (*Face Image Quality Variation Dataset*), by mal zahŕňať rôzne formy poškodení, ktoré sa vyskytujú v reálnych situáciách. FIQVD vychádza z CelebA[45] datasetu, ktorý sa často využíva na tréning a hodnotenie algoritmov počítačového videnia. Obsahuje viac ako 200 000 obrázkov tvárí s rôznymi atribútmi a vyniká podrobnou anotáciou jednotlivých snímok.

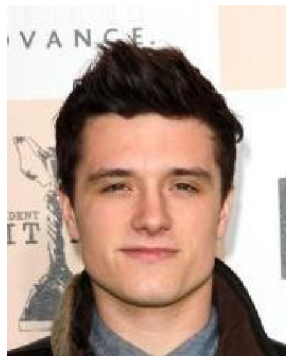


### 4.1.1 Degradacia snímok

Na zabezpečenie kvality a rozmanitosti datasetu je kľúčové mať dostatočný počet snímok pre každý typ poškodenia, na ktorý sa bude model trénovať. Tieto snímky boli vytvorené úpravou nepoškodených obrázkov z CelebA datasetu. FIQVD obsahuje obrázky rozdelené do 6 kategórií, ktoré zohľadňujú reálne situácie ovplyvňujúce kvalitu a identifikovateľnosť tváří v obraze.

#### Snímky bez zjavných poškodení

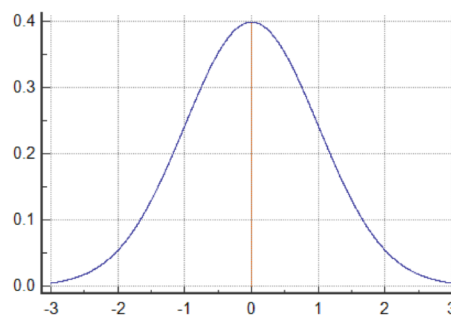
Tieto obrázky slúžia ako referenčný bod pre kvalitu a môžu pomôcť modelu pri klasifikácii snímok. Ich prítomnosť v datasete poskytuje modelu kontext a umožňuje mu efektívne rozlišovať medzi nepoškodenými a poškodenými snímkami. Príklad nepoškodenej (obr. 4.2) snímky je zobrazený nižšie.



Obr. 4.2: Snímka bez poškodení.

#### Rozmazané snímky

Dané snímky vznikli aplikovaním metódy gaussovho rozostrenia na obrázky. Táto operácia, podobná rámčekovému rozmazaniu (*box blurring*), je metódou dolnopriepustného filtra slúžiacia k rozmazaniu obrázkov [53]. Na vstupný obrázok aplikuje operáciu konvolúcie s gaussovským jadrom (*gaussian kernel*). Toto jadro je pomenované podľa gaussovej distribúcie (zvonovito tvarovaného pravdepodobnostného rozdelenia). Graf generickej gaussovej pravdepodobnostnej distribúcie so štandardnou odchýlkou  $\sigma = 1$  a priemerom v bode  $x = 0$  je znázornený na obrázku 4.3.



Obr. 4.3: Graf generického gaussovho pravdepodobnostného rozdelenia.

Funkciu na výpočet hodnoty jednotlivých prvkov dvojrozmerného gaussovho jadra môžeme definovať ako [10]:

$$G(i, j) = \frac{1}{2\pi\sigma^2} e^{-\frac{i^2+j^2}{2\sigma^2}} \quad (4.1)$$

Kde:  $G(i, j)$  je hodnota gaussovho jadra na súradniciach  $(i, j)$ ,  $\sigma$  je štandardná odchýlka gaussovho rozdelenia, ktorá určuje mieru rozmazania a  $(i, j)$  sú súradnice relatívne k stredu jadra.

Hodnoty  $G(i, j)$  predstavujú maticu, opisujúcu váhy susedných pixelov pri konvolučných operáciách. Na obrázku 4.4 môžeme vidieť diskkrétne hodnoty tohto jadra pri veľkosti matice  $3 \times 3$  a  $5 \times 5$ .

$1/16$	1	2	1
	2	4	2
	1	2	1

$1/273$	1	4	7	4	1
	4	16	26	16	4
	7	26	41	26	7
	4	16	26	16	4
	1	4	7	4	1

Obr. 4.4: Diskrétne aproximácia gaussovho jadra o veľkosti  $3 \times 3$  a  $5 \times 5$  [62].

Konvolúciou gaussovho jadra a pôvodného obrázku získame požadovaný obrázok deformovaný gausovým rozmazaním. Aplikáciu gaussovho rozmazania na jednom farebnom kanáli môžeme definovať:

$$I_{blurred}(x, y) = \sum_{i=-2k}^{2k} \sum_{j=-2k}^{2k} G(i, j) \cdot I(x + i, y + j) \quad (4.2)$$

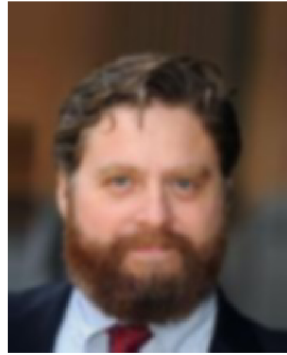
Kde:  $I_{blurred}(x, y)$  je hodnota pixelu daného farebného kanálu na súradniciach  $(x, y)$ ,  $k$  je veľkosť gaussovho jadra (nepárne číslo, ako napríklad: 3, 5, 7, a pod.) a  $I(x + i, y + j)$  je hodnota pixelu daného farebného kanálu na súradniciach  $(x + i, y + j)$  pôvodného obrázku

Pre každý pixel v každom farebnom kanáli sa pomocou Gaussovho jadra vypočíta vážený priemer susedných pixelov v rovnakom kanáli. Výsledná hodnota je priradená k zodpovedajúcemu umiestneniu pixelov v rozmazanom obrázku pre každý farebný kanál. Tento proces vyhladzuje šum a znižuje vysokofrekvenčné zložky každého farebného kanálu, čo vedie k rozmazanej verzii pôvodného farebného obrázku. Jedná sa o bežnú techniku vyhladzovania a rozmazávania obrázkov.

Porovnanie pôvodného (obr. 4.5a) a rozmazaného obrázka (obr. 4.5b) s veľkosťou gaussovho jadra  $i = 5, j = 5$  pomocou OpenCV knižnice [51] je zobrazené na nižšie.  $\sigma$  je vypočítaná na základe veľkosti jadra.



(a) Pôvodný obrázok



(b) Rozmazaný obrázok

Obr. 4.5: Obrázok poškodený rozmazaním.

### Snímky obsahujúce šum

Boli vytvorené pridaním gaussovho šumu. Ide o typ štatistického šumu, ktorý sa bežne vyskytuje pri spracovaní obrazu. Gaussov šum, často generovaný elektronickými komponentmi ako sú zosilovače a detektory, je bežným javom v digitálnych zobrazovacích zariadeniach. Tento šum vzniká vďaka prirodzeným javom, ako sú tepelné vibrácie atómov a charakter žiarenia teplých predmetov [52]. Je charakterizovaný svojím pravdepodobnostným rozdelením 4.3 a prejavuje sa ako náhodná zmena hodnôt pixelov, ktorá vedie k zrnitému vzhľadu obrázka. Tento jav môže mať za následok stratu detailu, čo sťažuje presnosť analýzy a interpretácie obrazu.

Vo farebných digitálnych obrázkoch gaussov šum narúša nielen hodnoty jasu (ako pri čiernobielych obrázkoch), ale aj hodnoty farebných kanálov, čo môže viesť k separátnemu skresleniu v rámci jednotlivých farieb (RGB). Gaussov šum je pre každý farebný kanál spojený s funkciou hustoty pravdepodobnosti, ktorú môžeme definovať ako [58]:

$$P(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(n-\mu)^2}{2\sigma^2}} \quad (4.3)$$

Kde:  $n$  predstavuje náhodný šum v konkrétnom farebnom kanáli,  $\mu$  je stredná hodnota gaussovho rozdelenia a  $\sigma$  je štandardnou odchýlkou, ktorá určuje intenzitu šumu.

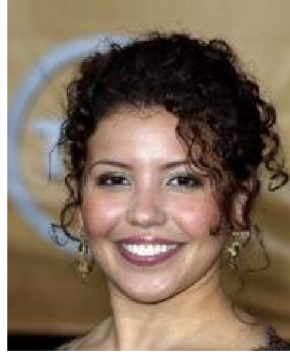
Pridanie gaussovho šumu zahŕňa prekrytie pôvodného obrázku náhodnými hodnotami získaných z gaussovho rozdelenia. Rovnica pre získanie hodnoty pixelu obrázku po pridaní gaussovho šumu na jednom farebnom kanáli môže byť vyjadrená:

$$I_{noisy}(x, y) = I(x, y) + n(x, y) \quad (4.4)$$

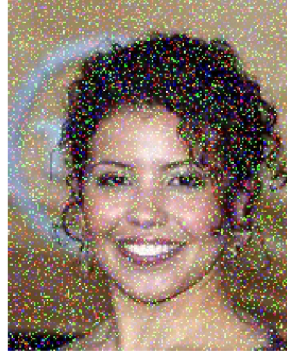
Kde:  $I_{noisy}(x, y)$  je hodnota výsledného pixelu na súradniciach  $(x, y)$ ,  $I(x, y)$  je hodnota pôvodného pixelu na súradniciach  $(x, y)$  a  $n(x, y)$  je hodnota gaussovho šumu na súradniciach  $(x, y)$ .

Vo farebných obrázkoch sa táto operácia vykonáva pre každý farebný kanál oddelene. To znamená že gaussov šum je pridávaný nezávisle od každého kanálu obrázku, čo môže viesť k rôznym úrovniam šumu v jednotlivých kanáloch.

Porovnanie pôvodného obrázku (obr. 4.6a) a obrázku so šumom (obr. 4.6b) s hodnotami  $\mu = 0$  a  $\sigma = 1$  pomocou knižnice OpenCV [51] je zobrazené nižšie.



(a) Pôvodný obrázok



(b) Obrázok so šumom

Obr. 4.6: Obrázok poškodený šumom.

### Snímky obsahujúce kompresné artefakty

Vznikli kompresiou a následnou dekompresiou obrázkov vo formáte JPEG. Štandardný proces JPEG kompresie zahŕňa: transformáciu obrazu z priestorovej domény do frekvenčnej oblasti pomocou diskkrétnej kosínovej transformácie (DCT), kvantovanie transformovaných koeficientov a kódovanie kvantovaných koeficientov.

DCT sa aplikuje nezávisle na každom farebnom kanáli (R, G, B) obrazu. Dvoj-rozmerná DCT operácia transformuje priestorovú doménu na jej reprezentáciu vo frekvenčnej doméne  $DCT(i, j)$  pre každý farebný kanál. Aplikáciu DCT pre jeden farebný kanál môžeme definovať ako [8]:

$$DCT(i, j) = \frac{1}{\sqrt{2N}} C(i) C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p(x, y) \cos \left[ \frac{(2x+1)i\pi}{2N} \right] \cos \left[ \frac{(2y+1)j\pi}{2N} \right] \quad (4.5)$$

Kde:  $N$  je veľkosť bloku (zvyčajne 8x8 pixelov),  $p(x, y)$  reprezentuje intenzitu pixelu obrázku na súradniciach  $(x, y)$ ,  $i$  a  $j$  predstavujú indexy priestorovej frekvencie (určujú konkrétne úrovne detailu) a  $C(i)$  a  $C(j)$  sú normalizačné faktory definované ako:

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{pre } u = 0, \\ 1, & \text{pre } u > 0. \end{cases} \quad (4.6)$$

Výstupom DCT je dvojrozmerná matica frekvenčných koeficientov. Každý frekvenčný koeficient  $DCT(i, j)$  sa kvantuje nezávisle pomocou kvantizačnej matice  $Q(i, j)$ , ktorá sa líši určením kvality kvantizácie (1 až 100). Hodnoty kvantizačnej matice môžu mať rozdielne hodnoty taktiež aj na základe farebného kanálu, na ktorom je operácia vykonávaná. Pri tomto procese nastáva delenie každého prvku v transformovanej matici korešpondujúcim prvkom v kvantizačnej matici, a následným zaokrúhlením na najbližšiu integer hodnotu. Znižuje sa počet bitov potrebných na uloženie celočíselnej hodnoty znížením presnosti celého čísla koeficientu, čiže sa redukujú menej užitočné dáta (nastáva kompresia). Kvantizáciu jedného farebného kanálu môžeme definovať ako [8]:

$$K(i, j) = \text{round} \left( \frac{DCT(i, j)}{Q(i, j)} \right) \quad (4.7)$$

Kde:  $Q(i, j)$  je kvantizačná matica, líšiaca sa určením hodnoty kvality a farebným kanálom,  $\text{round}()$  zaokrúhľuje na najbližšiu integer hodnotu.

Výsledkom kvantizácie je matica kvantovaných frekvenčných koeficientov  $K(i, j)$  pre každý farebný kanál. Dané matice môžu byť následne zakódované pomocou rôznych techník, ktoré určujú úroveň straty (alebo bezstratovosť) kompresie koeficientov. Účelom kompresie obrázkov je optimalizácia ukladania dát a zníženie veľkosti súborov, čo umožňuje efektívnejší prenos dát cez sieťové pripojenie a redukciu miesta, ktoré obrázok zaberá na úložisku.

Štandardný proces JPEG dekompresie zahŕňa inverzné operácie ako pri kompresii, akurát v opačnom poradí, teda: dekódovanie kvantovaných koeficientov, inverzná kvantizácia a inverzná diskretná kosínová transformácia (IDCT).

Pre získanie kvantovaných frekvenčných koeficientov  $K(i, j)$  sa všetky zakódované dáta dekódujú (spôsob dekódovania závisí od použitého zakódovania). K získaniu rekonštruovanej matice frekvenčných koeficientov  $R(i, j)$  sa vykoná proces dekvantizácie, ktorý môže byť matematicky definovaný ako [8]:

$$R(i, j) = Q(i, j) \times K(i, j) \quad (4.8)$$

Na rekonštruovanú maticu  $R(i, j)$  sa aplikuje IDCT zaokrúhlená k najbližšej integer hodnote. Pridaním polovice maximálnej hodnoty farebného kanálu (pri 8-bit formáte ide o hodnotu 128), vykompenzujeme posun (bias), ktorý nastal pri kvantizácií, a vzniká dekomprimovaný obraz. Výpočet pixelov dekomprimovaného obrazu pre jeden farebný kanál (8-bit) môžeme definovať ako [8]:

$$p(x, y) = \text{round} \left( \frac{1}{\sqrt{2N}} C(i) C(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} DCT(i, j) \cos \left[ \frac{(2x+1)i\pi}{2N} \right] \cos \left[ \frac{(2y+1)j\pi}{2N} \right] \right) + 128 \quad (4.9)$$

Kde hodnoty premenných majú rovnaký význam ako pri DCT.

Porovnanie pôvodného (obr. 4.7a) a komprimovaného, a následne dekomprimovaného obrazu (obr. 4.7b) s hodnotou kvality 20 pomocou OpenCV knižnice je zobrazené nižšie.



(a) Pôvodný obrázok



(b) Obrázok so šumom

Obr. 4.7: Obrázok poškodený kompresiou.



## Snímky so zakrytou hornou časťou tváre

Vznikli prekrytím periokulárnej časti tváre na obrázkoch. Transformáciou obrázka na čier-nobiely je možné detegovať tvár pomocou HOG detektoru [16]. Hoci ide o metódu detekcie tváre, ktorá je považovaná za zastaranú, nedokáže úplne presne identifikovať tváre na obrázkoch s výraznými rozdielmi v osvetlení a nezabezpečuje spoľahlivú detekciu pri vysokých stupňoch rotácie hlavy, čo je želané pri nastávajúcej degradácii pomocou oklúzií z dôvodov popísaných nižšie. Po lokalizácii tváre na obrázku umožňuje shape prediktor *68 landmarks* identifikovať súradnice bodov potrebných na zakrytie v oblasti očí, ako sú stred nosa (spodná hranica), vonkajšie kútiky očí (ľavá a pravá hranica) a vyššie položené obočie (horná hranica). Tieto body (obr. 4.8a) následne slúžia na určenie pozície obdĺžnikového útvaru, ktorý definuje oblasť prekrytia (obr. 4.8b), do ktorej môže byť pridaný objekt určený na prekrytie, napríklad slnečné okuliare.



(a) Obrázok so zvýraznenými hraničnými bodmi

(b) Obrázok so zakrytím v oblasti záujmu

Obr. 4.8: Hraničné body hornej časti tváre, a oblasti, ktoré ohraničujú.

Pre zachovanie oblastí v rámci ROI, ktoré nebudú zakryté okuliarmi, sa vyžadujú transparentné obrázky (napríklad PNG formát). Následne sa obrázky zmenšia tak, aby zodpovedali veľkosti ROI. Táto zmena veľkosti obrázkov je realizovaná pomocou bikubickej interpolácie, ktorá je technikou používanou pri zmenách veľkosti obrázkov počítaním hodnôt pixelov v obrázku so zmenenými rozmermi pomocou váženého priemeru susedných pixelov v pôvodnom obrázku, čo prispieva k zachovaniu detailov z pôvodného obrázku. Funkciu bikubickej interpolácie možno matematicky vyjadriť ako [75]:

$$f(x) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} \cdot x^i \cdot x^j \quad (4.10)$$

Kde:  $x$  a  $y$  sú horizontálne a vertikálne vzdialenosti od cieľového pixelu k najbližšiemu bodu v matici pôvodného obrazu a  $a_{ij}$  sú interpolačné koeficienty určené jadrom kubickej konvolúcie.

Presné koeficienty a detaily implementácie bikubickej interpolácie sa líšia v závislosti od konkrétnej softvérovej knižnice a použitého algoritmu. Napriek tomu základná myšlienka ostáva rovnaká. Hodnota každého pixelu v zmenenom obrázku sa vypočíta vzhľadom na vážený priemer susedných pixelov na základe jadra kubickej konvolúcie.

Extrahovaním kanálu priehľadnosti (alpha channel) z obrázka okuliarov so zmenenou veľkosťou sa získajú informácie o priehľadnosti jednotlivých pixelov. Normalizáciou týchto

hodnôt a dopočítaním komplementárnej priehľadnosti pôvodného obrázka sa zabezpečí, že pixely pôvodného obrázka, ktoré budú zakryté okuliarmi, budú transparentné, zatiaľ čo ostatné zostanú nezmenené. Výsledný obrázok so zakrytím sa získa operáciou miešania hodnôt pixelov v každom farebnom kanáli. Operáciu miešania pre jeden farebný kanál môžeme definovať ako:

$$I_{blended}(x, y) = (A_g(x, y) \times I_g(x, y) + (A_{ROI}(x, y) \times I_{ROI}(x, y)) \quad (4.11)$$

Kde:  $I_{blended}(x, y)$  je hodnota pixelu jedného farebného kanálu výsledného obrázku,  $A_g(x, y) \times I_g(x, y)$  je hodnota pixelu okuliarov na súradniciach  $(x, y)$  a  $A_{ROI}(x, y) \times I_{ROI}(x, y)$  je hodnota pixelu pôvodného obrázku na súradniciach  $(x, y)$ .

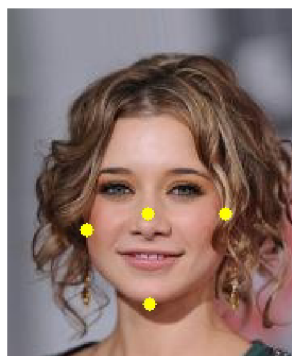
Treba avšak poznamenať, že pri tomto postupe sa nepredpokladá prívětivé naklonenie a rotácia hlavy, čo má za následok menej dôveryhodné snímky, ak nie je snímaná osoba otočená čelom ku kamere. Z tohto dôvodu bola pre účely detekcie použitá zastaralejšia metóda detekcie pomocou HOG. Porovnanie pôvodného obrázku (obr. 4.9a) s obrázkom s oklúziou (obr. 4.9b) je zobrazené nižšie.



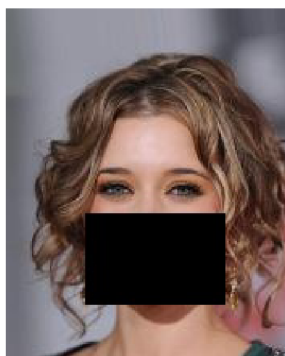
Obr. 4.9: Obrázok poškodený zakrytím hornej časti tváre.

### **Snímky so zakrytou spodnou časťou tváre**

Dané obrázky boli vytvorené podobným spôsobom ako snímky so zakrytou hornou časťou tváre. Jediný rozdiel spočíva v hraničných bodoch identifikovaných pomocou shape prediktora [35] a použitých obrázkov pri oklúzií. Pokiaľ ide o hraničné body tvoriace oblasť záujmu (ROI), zahŕňajú spodok brady (dolnú hranicu), stred nosa (hornú hranicu) a vonkajšie okraje čeluste (pravú a ľavú hranicu). Tieto body (obr. 4.10a) a oblasť prekrytia (obr. 4.10b) v pôvodnom obrázku sú zobrazené nižšie.



(a) Obrázok so zvýraznenými hraničnými bodmi



(b) Obrázok so zakrytím v oblasti záujmu

Obr. 4.10: Hraničné body spodnej časti tváre, a oblasti, ktoré ohraničujú.

Na zakrytie boli použité voľne dostupné PNG snímky masiek. Porovnanie pôvodného (obr. 4.11a) a výsledného obrázka so zakrytím (obr. 4.11b) je prezentované na obrázku nižšie.



(a) Pôvodný obrázok



(b) Obrázok so zakrytím

Obr. 4.11: Obrázok poškodený zakrytím spodnej časti tváre.

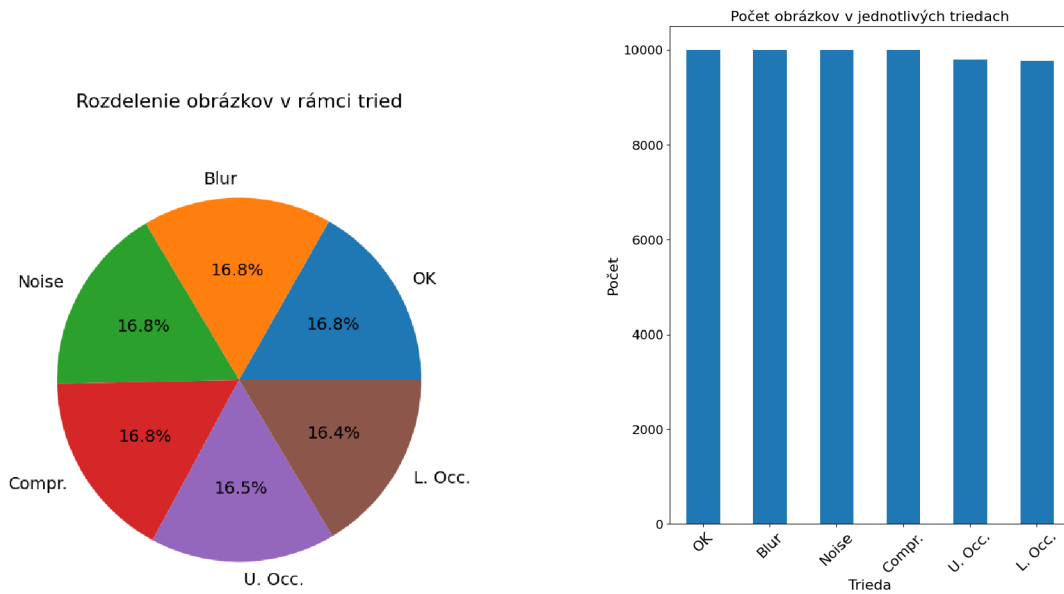
#### 4.1.2 Rozdelenie snímok v datasete

Počet snímok v jednotlivých kategóriách je dôležitým faktorom pre rovnomerné a efektívne tréovanie modelov strojového učenia. Zabezpečuje to, že modely majú dostatočné množstvo tréovacích príkladov pre každý typ poškodenia a sú schopné naučiť sa relevantné vzory a charakteristiky.

Správne rozdelenie snímok do kategórií v rámci FIQVD zabezpečuje, že tréovacie dáta sú vyvážené a modely majú adekvátnu reprezentáciu každej kategórie poškodenia. Týmto spôsobom sa minimalizuje riziko skreslenia výsledkov a zlepšuje sa schopnosť modelov generalizovať na nové, neznáme snímky.

Detaily o rozdelení a počte snímok v tomto datasete sú zobrazené nižšie (obr. 4.12).





Obr. 4.12: Rozdelenie snímok v FIQVD.

Pre tréning, validačné a testovacie účely bol FIQVD rozdelený v pomere 6 : 2 : 2, čo zodpovedá 60 % tréningových, 20 % validačných a 20 % testovacích dát. Takéto rozdelenie umožňuje primerané množstvo dát pre tréning modelu, zatiaľ čo validačná a testovacia sada umožňujú overiť jeho výkon a generalizáciu na nové, nevidené snímky. Rozdelenie datasetu v tomto pomere je štandardnou praxou pri strojovom učení, ktorá pomáha zabrániť pretrénovaniu a poskytuje spoľahlivejšie vyhodnotenie modelu.

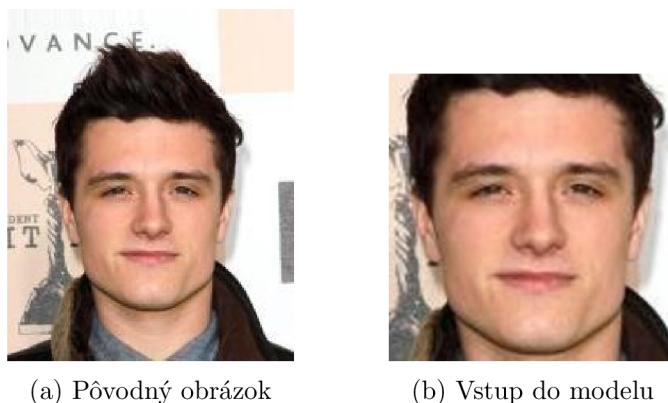
### 4.1.3 Predspracovanie snímok

Proces predspracovania obrazu v kontexte rozpoznávacích úloh slúži primárne k separácii nepodstatných častí pôvodného obrázku od častí, na ktorých má byť vykonávané rozpoznávanie [41]. Predspracovanie môže zahŕňať širokú škálu techník slúžiacich k optimalizácii kvality a vhodnosti obrazových dát, alebo urýchleniu algoritmu spracovania obrazu.

V kontexte konvolučných neurónových sietí (CNN) je predspracovanie obrazu často spojené s transformáciou obrazu do formátu vhodného pre model. Tento proces je kľúčový, pretože kvalita vstupných dát výrazne ovplyvňuje výsledky analýzy alebo klasifikácie obrazu. Korektné predspracovanie obrazových dát môže viesť k zlepšeniu presnosti a úspešnosti modelov strojového učenia.

Aby obrázky z FIQVD datasetu odpovedali vyžadovanému vstupu modelu, je potrebné vykonať zarovnanie na štvorec vzhľadom na tvár, a následne upraviť rozmery tohto štvorca na  $224 \times 224$  pixelov. Zarovnanie bolo vykonané pomocou RetinaFace detektoru z knižnice FaceAIKit [24]. Pre zachovanie podstatných detailov obrázku (vrátane šumu, rozmazania a kompresných artefaktov) bola zmena veľkosti vykonaná bikubicou interpoláciou z knižnice OpenCV [51] popísanou vyššie.

Príklad pôvodného obrázku (obr. 4.13a) a obrázku spracovaného na vstup do modelu neurónovej siete (obr. 4.13b) je zobrazený nižšie.



Obr. 4.13: Spracovanie obrázka na vstup do modelu.

Je potrebné spomenúť, ako sa správajú poškodenia spôsobené rozmazaním, šumom a kompresiou, pri snímkach značne iných rozmerov, ako pri tréningových snímkach (napr.  $3000 \times 3000$  alebo  $64 \times 64$  pixelov). Zmenou rozmerov štvorcovej oblasti na rozmery odpovedajúce vstupu modelu ( $224 \times 224$ ), nastane jav nazývaný *downscale* alebo *upscale*.

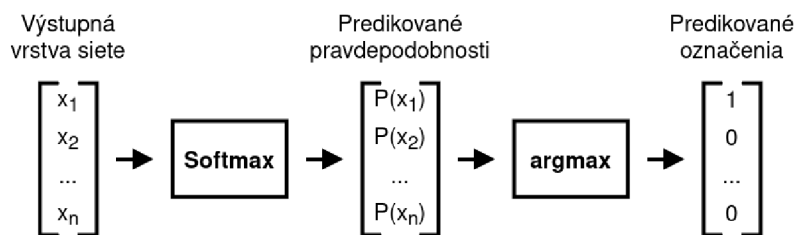
*Downscale* predstavuje zmenšenie rozmerov obrazových dát oproti pôvodnému. Počas tohto javu sa degradácie rozmazania, šumu a kompresie stanú (takmer) nedetekovateľnými modelom implementovaným v tejto práci, pretože *downscaling* potláča vysokofrekvenčné zložky (napr. šum a kompresné artefakty) a nízkofrekvenčné zložky (rozmazanie) sú rozložené vo výslednom obraze nižších rozmerov, čím sa navyšuje ostrosť.

Naopak, *upscale* predstavuje navýšenie rozmerov obrazových dát. Počas tohto javu sú degradácie obrazu výraznejšie, pretože na vyplnenie oblasti medzi pôvodnými pixelmi sa generujú nové hodnoty pixelov. Tento proces môže viesť k zvýšeniu degradácií, pretože existujúce degradácie sa rozšíria na viac pixelov.

## 4.2 Implementácia architektúry ResNet

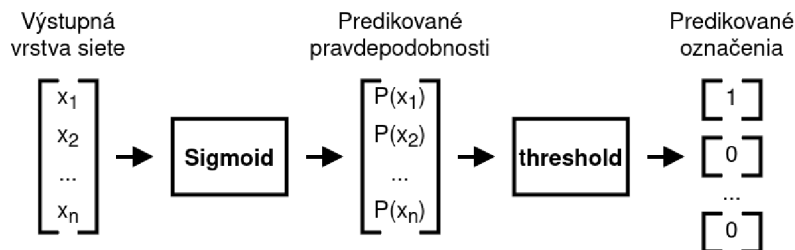
Modely implementované v tejto práci sú ResNet-18 a ResNet-50. Na rozdiel od pôvodnej architektúry [29], kde mala posledná plne prepojená vrstva 1000 výstupov, bol počet výstupov redukovaný na počet kategórií snímok v FIQVD datasete (čiže 6). Ďalším rozdielom je aktivačná funkcia (alebo jej absencia) nasledujúca za touto vrstvou.

Pri tréovaní siete pomocou stratovej funkcie Cross-Entropy Loss bola posledná vrstva lineárna, pretože daná funkcia očakáva nenormalizovaný vstup<sup>1</sup>. Daná funkcia aplikuje aktivačnú funkciu Softmax na výstup siete k získaniu pravdepodobností jednotlivých tried. Trieda s najvyššou pravdepodobnosťou je označená ako predikovaná trieda (obr. 4.14).



Obr. 4.14: Predikcia pomocou stratovej funkcie Cross-Entropy Loss.

Pri tréovaní pomocou stratovej funkcie Mean Squared Error (MSE) bola posledná vrstva zakončená aktivačnou funkciou Sigmoid, taktiež nazývanú ako logistická funkcia. Napriek tomu, že MSE je bežne používaná najmä pri regresných úlohách, jej použitie je opodstatnené ak vstup môže patriť do viacerých tried (napríklad snímky obsahujúce viacero degradácií). Ako predikované triedy boli označené triedy, ktoré prekročili pravdepodobnostný prah ( $threshold^2$ ) (obr. 4.15).



Obr. 4.15: Predikcia pomocou stratovej funkcie MSE.

Ako optimalizačný algoritmus bol použitý Adam z dôvodov opísaných v teoretickej časti tejto práce 2.2.2.

<sup>1</sup>Implementácia Cross-Entropy Loss v knižnici PyTorch je ekvivalentná aplikácií funkcie LogSoftmax nasledovanú NLLLoss (Negative Log Likelihood Loss) na výstup siete.

<sup>2</sup>Prah pravdepodobností všetkých tried bol nastavený na 0.5, ak nastala situácia, kedy žiadna predikovaná pravdepodobnosť neprekročila tento prah, ako predikovaná trieda bola označená tá, ktorá dosahovala najvyššiu pravdepodobnosť, aby sa zabránilo tomu, že model nepredikoval žiadnu triedu.

### 4.3 Výsledná aplikácia

Aplikácia bola implementovaná pomocou Flask frameworku. Využíva natrénovaný model k vyhodnoteniu jednotlivých pravdepodobností kategórie snímky a výsledného skóre kvality.


Užívateľ predáva snímku tváre prostredníctvom jednoduchého webového formulára. Ak ide o obrázok iného formátu ako JPEG, vstupná snímka je do tohto formátu konvertovaná. Nevýhnutnosť formátu JPEG plynie z počtu farebných kanálov, ktoré na vstupe vyžaduje model hlbokoj neurónovej siete (čiže 3). Na snímke požadovaného formátu sa následne vykoná zarovnanie tváre na štvorec (popísané v sekcii 4.1.3). Zarovnaná snímka dosahuje všetky parametre potrebné na jej vstup do modelu ( $3 \times 224 \times 224$ ).

Model predikuje pravdepodobnosti jednotlivých tried na základe vlastností danej fotografie, a pomocou týchto predikcií sa vyhodnocuje výsledné skóre kvality. Po vyhodnotení všetkých potrebných parametrov sa užívateľovi zobrazí stránka so zarovnanou snímkou, tabuľkou pravdepodobností tried, výslednou klasifikáciou snímky a skóre kvality snímky.

Snímka výslednej aplikácie je zobrazená na obr. 4.16.

## Upload a Face Image

Browse...No file selected.Upload

### Predicted Probabilities:

OK	Blurr	Noise	Compression	Upper Face Occlusion	Lower Face Occlusion
0.01	0.03	0.05	0.05	0.84	0.02

The uploaded image **has Upper Face Occlusion**

Quality Score: **0.549**

Obr. 4.16: Snímka výslednej aplikácie.

### 4.4 Skóre kvality

Natrénovaný model predikuje pravdepodobnosti pre každú triedu, do ktorej vstupná snímka môže patriť. Výsledné skóre kvality  $Q$  je vypočítané na základe predpovedaných pravdepodobností a váh im priradeným (tab. 4.1).

Tabuľka 4.1: Tabuľka váh jednotlivých tried.

OK	Rozmazanie	Šum	Kompresia	Horné Zakrytie	Spodné Zakrytie
1	0.725	0.585	0.737	0.525	0.535

Váhy jednotlivých pravdepodobností tried boli určené na základe priemerovanej podobnosti pomocou euklidovskej vzdialenosti (norma  $L_2$ ), medzi pôvodnými snímkami z datasetu CelebA a rovnakými, akurát degradovanými snímkami z FIQVD. Výpočet podobnosti tváří zo snímkov môžeme jednoducho vypočítať ako:

$$S = \frac{1}{1 + d_{L_2}} \quad (4.12)$$

Kde:  $S$  je podobnosť medzi tvármi a  $d_{L_2}$  je euklidovská vzdialenosť medzi tvármi.

Euklidovská vzdialenosť  $d_{L_2}$  bola vypočítaná pomocou algoritmu rozpoznávania tváří ArcFace. Tento algoritmus je podrobnejšie popísaný v sekcii 2.4.

K vyjadreniu výsledného skóre kvality  $Q$  je potrebné pracovať s normalizovanými výstupnými pravdepodobnosťami modelu ( $P_i$ ) a váhami jednotlivých tried ( $W_i$ ). Normalizáciu výstupov pravdepodobností pri modeli natrénovanom pomocou stratovej funkcie Cross-Entropy Loss dosiahneme aktivačnou funkciou Softmax. Pri modeli natrénovanom pomocou stratovej funkcie MSE dosiahneme normalizáciu vydelením všetkých pravdepodobností sumou hodnôt pravdepodobností. Normalizované pravdepodobnosti nám potom dávajú súčet rovný 1, vďaka čomu môžeme jednoducho určiť výsledné skóre kvality  $Q$  definované ako:

$$Q = \sum_{i=1}^n P_i \times W_i \quad (4.13)$$

Kde:  $P_i$  je normalizovaná pravdepodobnosť, že snímka patrí triede  $i$ ,  $W_i$  je váha priradená triede  $i$  a  $n$  je počet tried

Dané skóre vie slúžiť ako orientačná kvantifikácia kvality pri poškodených snímkach, je však potrebné poznamenať, že sa počas jeho výpočtu neberú do úvahy určité kritické faktory, ako sú rotácia tváre a nerovnomerné osvetlenie. Taktiež je dôležité podotknúť, že ak model predikuje napríklad jedno to isté poškodenie s takmer stopercentnou pravdepodobnosťou na dvoch rozdielnych fotkách (napr. stopercentná pravdepodobnosť rozmazania), tak model vyhodnotí kvalitu daných snímkov rovnako, pretože neurčuje priamo konkrétne miery poškodení, iba pravdepodobnosti ich výskytov.

## Kapitola 5

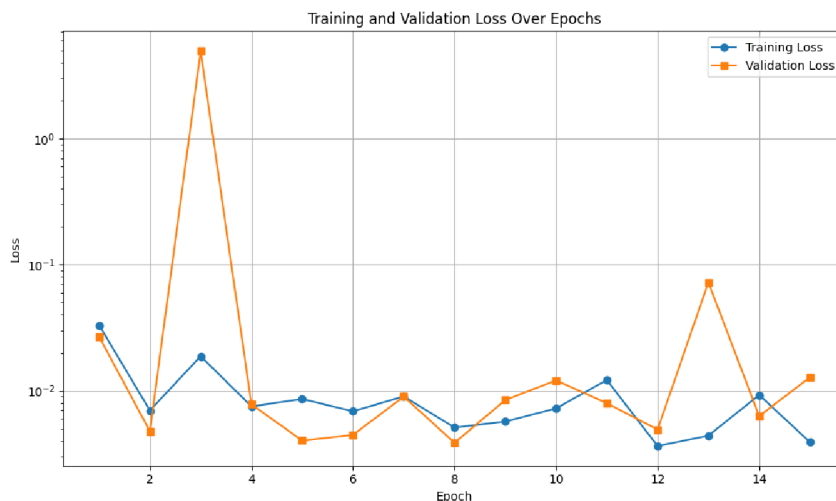
# Experimenty

Modely konvolučných neurónových sietí Resnet-18 a Resnet-50 boli tréňované pomocou stratových funkcií *Cross-Entropy Loss* a MSE pomocou variácií hyperparametrov:

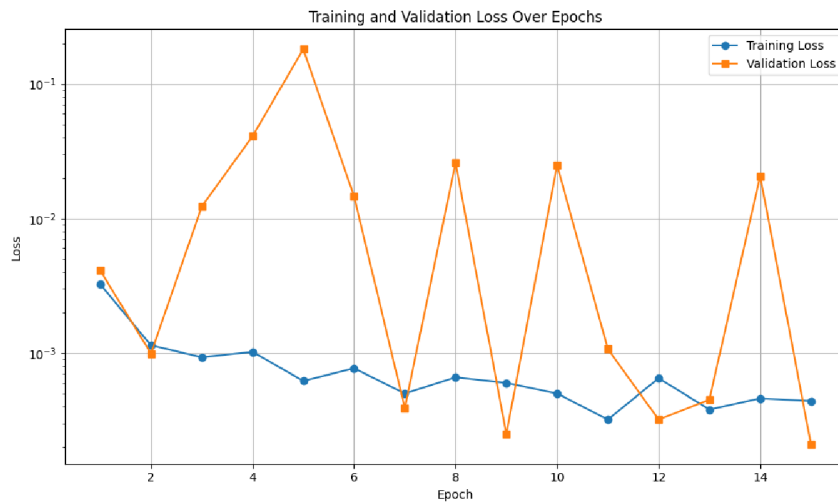
- $num\_epochs = \{5, 15, 30\}$
- $batch\_size = \{32, 64\}$
- $learning\_rate = \{1 \cdot 10^{-3}, 5 \cdot 10^{-4}\}$
- $weight\_decay = \{0, 1 \cdot 10^{-4}, 1 \cdot 10^{-5}\}$

Na základe hodnotiacich metrick popísaných v sekcii 5.1 sú v následných grafoch zobrazené tréňovacie a validačné straty modelov, ktoré dosahovali najlepšie výsledky.

Priebeh logaritmizovaných tréňovacích a validačných strát modelov ResNet-50 pri:  $num\_epochs = 15$ ,  $batch\_size = 64$ ,  $learning\_rate = 1 \cdot 10^{-3}$  a  $weight\_decay = 1 \cdot 10^{-5}$ , je zobrazený na obrázku 5.1 pre stratovú funkciu Cross-Entropy Loss, a na obrázku 5.2 pre stratovú funkciu MSE. Logaritmizovanie hodnôt pomáha pri vyhladení extrémne vysokých hodnôt údajov (hroty v grafe), čo robí výsledný graf čitateľnejší.

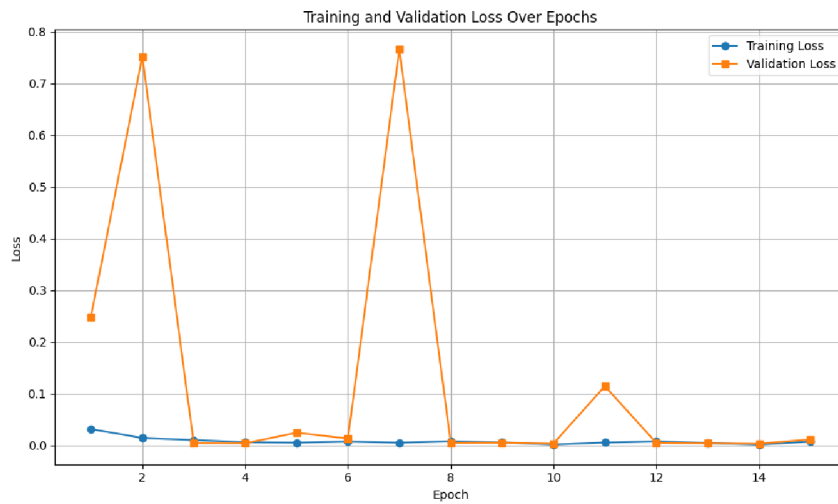


Obr. 5.1: Logaritmizované stratové hodnoty funkcie Cross-Entropy-Loss modelu ResNet-50.

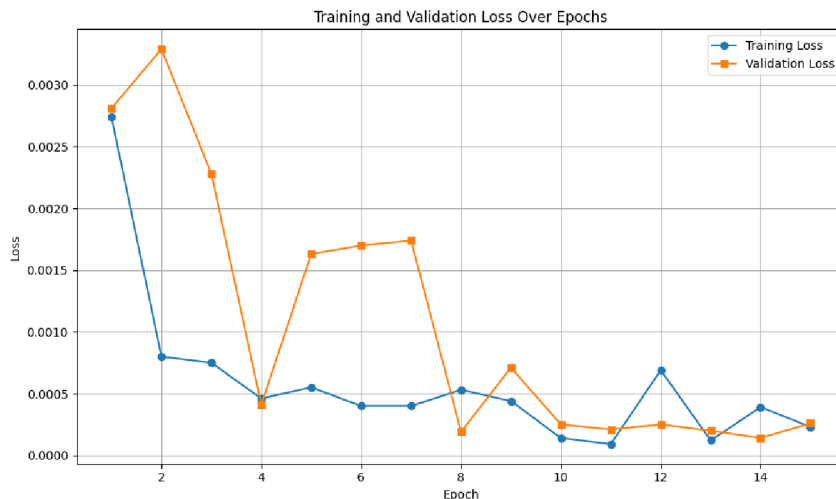


Obr. 5.2: Logaritmizované Stratové hodnoty funkcie MSE modelu ResNet-50.

Priebeh tréningových a validačných strát modelov ResNet-18 pri:  $num\_epochs = 15$ ,  $batch\_size = 64$ ,  $learning\_rate = 1 \cdot 10^{-3}$  a  $weight\_decay = 0$ , je zobrazený na obrázku 5.3 pre stratovú funkciu Cross-Entropy Loss, a na obrázku 5.4 pre stratovú funkciu MSE.



Obr. 5.3: Stratové hodnoty funkcie Cross-Entropy-Loss modelu ResNet-18.



Obr. 5.4: Stratové hodnoty funkcie MSE modelu ResNet-18.

Prudké nárasty hodnôt validačných strát počas tréningu modelu označujú obdobia, kedy sa výkon modelu na validačných dátach náhle zhoršil. Tieto výkyvy môžu nastať z rôznych dôvodov, ako napríklad pretrénovanie modelu, náhle zmeny v hodnotách rýchlosti učenia (*learning\_rate*), alebo nestabilita modelu spôsobená samotnou architektúrou či inicializačným nastavením váh modelu.

## 5.1 Hodnotiace metriky

Na základe predikcie modelu a skutočného označenia danej vzorky môžeme vzorky rozdeliť nasledovne (obr. 5.5):

		Predikovaná trieda	
		Pozitívna	Negatívna
Skutočná trieda	Pozitívna	Skutočné pozitíva (TP)	Falošné negatíva (FN)
	Negatívna	Falošné pozitíva (FP)	Skutočné negatíva (TN)

Obr. 5.5: Rozdelenie vzoriek podľa ich predikcie a skutočného označenia [3].

Na základe tohto rozdelenia vzoriek vieme následne vypočítať potrebné metriky, ktoré nám poskytnú kvantitatívny pohľad na to, ako dobre je model schopný klasifikovať dáta. Ide o metriky [3]:



$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (5.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (5.3)$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5.4)$$

Tieto metriky nám poskytujú dôležité informácie o výkone modelu počas klasifikácie. *Accuracy* vyjadruje percentuálnu úspešnosť modelu vo všeobecnosti. *Precision* hovorí, aký podiel pozitívnych identifikácií modelu je skutočne správnych. *Recall* meria, aký podiel skutočných pozitívnych príkladov je model schopný identifikovať. *F1-score* je harmonickým priemerom hodnôt *Precision* a *Recall*, ktorý poskytuje vyvážený pohľad na výkonnosť modelu. Tieto metriky sú dôležité nástroje pri hodnotení výkonnosti klasifikačných modelov.

Počas hodnotenia úspešnosti predikcií všetkých modelov ResNet-18 a Resnet-50 na testovacej časti FIQVD dosahoval *Accuracy* nad 99% a zvyšné metriky jednotlivých tried dosahovali rovnaké hodnoty (tab. 5.1).

Tabuľka 5.1: Tabuľka hodnotiacich metrick pri testovaní na FIQVD.

	Precision	Recall	F1-score	Počet vzoriek
Bez poškodenia	1.00	1.00	1.00	2015
Rozmazanie	1.00	1.00	1.00	1998
Šum	1.00	1.00	1.00	1948
Kompresia	1.00	1.00	1.00	2005
Horné zakrytie	1.00	1.00	1.00	2039
Spodné zakrytie	1.00	1.00	1.00	1911

Na základe uvedených metrick by bolo možné vyvodit, že predikcie modelov sú takmer bezchybné. Avšak existuje niekoľko dôvodov, prečo tabuľka 5.1 môže byť (a mala by byť) považovaná za nedôveryhodnú. Prvým z nich môže byť tzv. únik dát, čo znamená, že dáta, ktoré by nemali byť použité v procese tréovania, sú neúmyselne zahrnuté. Tento únik by mohol viesť k tomu, že model sa naučí rozpoznávať dané dáta už v priebehu tréovania, čo následne vedie k chybným záverom. Avšak v prípade popísaného v tejto práci únik dát nenastal, pretože nenastal prekryv medzi snímkami v tréovacej a testovacej časti FIQVD. Z toho vyplýva, že počas testovania bol model hodnotený len na snímkach, na ktorých nebol natréovaný.

Ďalším pravdepodobným dôvodom môže byť skutočnosť, že snímky v FIQVD sú veľmi podobné, a modely sa naučili rozpoznávať len určité charakteristiky, ktoré sa v tejto dátovej sade vyskytujú. Napríklad, rozmazanie sa vytvorilo iba pomocou gaussovho rozostrenia rovnakou veľkosťou gaussovho jadra, šum bol aplikovaný iba s určitou intenzitou gaussovho rozloženia, kompresia mala konštantnú hodnotu a oklúzie boli vykonávané pomocou malého počtu rôznych objektov. Z daných skutočností môže vyplývať, že nemusela nastat dostatočná variácia medzi poškodeniami jednotlivých snímkov.

Ďalším faktorom, ktorý by mohol spôsobiť nedostatočnú generalizáciu modelu v identifikácii jednotlivých poškodení, je to, že všetky snímky FIQVD boli vytvorené zo snímkov

CelebA datasetu, kedy ide o snímky rovnakého rozlíšenia, a model teda nemusel byť schopný detegovať degradácie pri iných rozlíšeniach.

Z týchto skutočností vyplýva, že je potrebné výsledný model otestovať na iných snímkach ako na testovacej časti FIQVD. Dané testovacie snímky pozostávajú zo snímkov ROF datasetu[17] pre oklúzie hornej a spodnej časti tváre, zvyšné snímky boli prevzaté z CelebA-HQ resized ( $256 \times 256$ ) datasetu[50], kde nepoškodené snímky sú neupravené snímky tohto datasetu. Snímky obsahujúce degradácie rozmazania, šumu a kompresie boli vytvorené úpravou pôvodných snímkov pri rôznych parametroch určujúcich degradáciu. Pre rozmazané snímky ide o veľkosti konvolučných jadier  $5 \times 5$ ,  $7 \times 7$  a  $9 \times 9$ . Pre snímky obsahujúce šum ide o intenzitu šumu  $\sigma = 0.2$ ,  $0.6$  a  $1.0$ . Pre komprimované snímky sa jedná o hodnotu kvality 10, 30 a 50.

Počas hodnotenia úspešnosti predikcií modelov pri empirickom testovaní na 300 snímkach boli hodnoty metrických spoľahlivosti modelov nasledovné:

- *Accuracy* = 0.74 a tab. 5.2, pre ResNet-50 trénovaného pomocou CEL
- *Accuracy* = 0.6366 a tab. 5.3 pre ResNet-50 trénovaného pomocou MSE
- *Accuracy* = 0.6366 a tab. 5.4 pre ResNet-18 trénovaného pomocou CEL
- *Accuracy* = 0.71 a tab. 5.5 pre ResNet-18 trénovaného pomocou MSE

Tabuľka 5.2: Tabuľka hodnotiacich metrických pri empirickom testovaní Resnet-50 trénovaného pomocou stratovej funkcie Cross-Entropy Loss.

	Precision	Recall	F1-score	Počet vzoriek
Bez poškodenia	0.43	1.00	0.60	50
Rozmazanie	1.00	0.66	0.80	50
Šum	1.00	0.66	0.80	50
Kompresia	0.98	0.92	0.95	50
Horné zakrytie	1.00	0.34	0.51	50
Spodné zakrytie	0.80	0.86	0.83	50

Tabuľka 5.3: Tabuľka hodnotiacich metrických pri empirickom testovaní Resnet-50 trénovaného pomocou stratovej funkcie MSE.

	Precision	Recall	F1-score	Počet vzoriek
Bez poškodenia	0.34	1.00	0.51	50
Rozmazanie	1.00	0.54	0.70	50
Šum	1.00	0.66	0.80	50
Kompresia	1.00	0.66	0.80	50
Horné zakrytie	1.00	0.20	0.33	50
Spodné zakrytie	0.76	0.76	0.76	50

Tabuľka 5.4: Tabuľka hodnotiacich metrík pri empirickom testovaní Resnet-18 trénovaného pomocou stratovej funkcie Cross-Entropy Loss.

	Precision	Recall	F1-score	Počet vzoriek
Bez poškodenia	0.33	1.00	0.49	50
Rozmazanie	0.86	0.50	0.63	50
Šum	1.00	0.54	0.70	50
Kompresia	0.96	0.44	0.60	50
Horné zakrytie	1.00	0.46	0.63	50
Spodné zakrytie	0.98	0.88	0.93	50

Tabuľka 5.5: Tabuľka hodnotiacich metrík pri empirickom testovaní Resnet-18 trénovaného pomocou stratovej funkcie MSE.

	Precision	Recall	F1-score	Počet vzoriek
Bez poškodenia	0.36	1.00	0.53	50
Rozmazanie	1.00	0.54	0.70	50
Šum	1.00	0.66	0.80	50
Kompresia	1.00	0.44	0.61	50
Horné zakrytie	1.00	0.78	0.88	50
Spodné zakrytie	1.00	0.88	0.94	50

Z týchto tabuliek vyplýva, že modely architektúr ResNet-50 mali najväčšie problémy pri detekcii degradácie snímok obsahujúcich zakrytie hornej časti tváre a rozmazaných snímok, zatiaľ čo modely ResNet-18 mali najväčší problém s detegovaním kompresie u snímok. Modely teda generalizujú vstupné dáta iba čiastočne, a pri tréovaní nastal *overfitting*.

Zaujímavým poznatkom je, že všetky modely dosahovali hodnotu  $Recall = 1$  u snímok bez poškodení, z čoho vyplýva, že všetky snímky, ktoré neboli poškodené, boli označené správne.

Ďalším poznatkom je, že posledný spomínaný model (tab. 5.5) síce nedosahoval najvyššiu hodnotu  $Accuracy$ , ale dosahoval hodnotu  $Precision = 1.0$  u každej degradácie, čo znamená, že všetky snímky, ktoré označil ako degradované, daným spôsobom degradované skutočne boli.

## 5.2 Zmeny v datasete

Prvé tréovania boli vykonávané na snímkach, kde boli črty tváre zakryté pomocou obdĺžnikových útvarov. Výplň útvaru bola buď jednoliata čierna farba, alebo biely šum odvodený od priemernej farby v pôvodnej oblasti, ktorá bola zakrytá. Modely natréované týmto spôsobom však nedosahovali uspokojujúce výsledky, a nedokázali detegovať oklúzie pri skutočných scénaroch (napríklad zakrytie očí pomocou slnečných okuliarov), pretože obdĺžnikové útvary boli príliš rozdielne od skutočných objektov, ktorými nastala oklúzia kľúčových črt.

Pri ďalšej tréovacej iterácii bol FIQVD pozmenený tak, aby boli zakrytia vykonané pomocou sady skutočných obrázkov slnečných okuliarov. Zakrytia spodnej časti tváre boli snímky prevzaté z MaskedFace-Net[7] datasetu. Pri testovaní modelov natréovaných týmto spôsobom však nastali problémy pri detekcii iných masiek, ako tých ktoré boli súčasťou

prevzatého datasetu (modré COVID-19 masky). Ďalším nedostatkom daného datasetu bolo, že obsahoval veľký počet snímok obsahujúcich prekrytie vrchnej aj spodnej časti tváre, čo mohlo viesť k nesprávnemu trénovaniu, pretože vzorky označené iba spodným zakrytím mali taktiež zakrytie periokulárnej oblasti, čo nebolo zohľadnené pri označovaní dátovej sady.

Tento problém bol odstránený pridaním oklúzií spodnej časti tváre rovnakým spôsobom, ako pri zakrytí hornej časti tváre. Nevýhodou oproti MaskedFace-Net datasetu je, že masky nie sú správne zrotované, a model nie je robustný voči variáciám v póze tváre pri zakrytiach kľúčových častí tváre.

Ostatné formy degradácie boli upravené variáciou parametrov rozmazaní (rôzne veľkosti gaussovho jadra), šume (rôzne intenzity) a kompresíí (rôzne miery degradácie). Trénovanie na týchto dátach spôsobilo nestabilitu modelu a nijak kladne neovplyvnilo jeho výkonnosť. Zmeny parametrov degradácií v datasete neboli zachované pre ďalšie trénovania.

## Kapitola 6

### Záver

Presnosť algoritmov pre rozpoznávanie tváre je výrazne závislá na kvalite vstupných snímok. Výsledkom práce je aplikácia využívajúca model hlbokoj neurónovej siete ResNet na vyhodnotenie kvality snímok tváre. Tento model analyzuje pravdepodobnosti výskytu rôznych typov poškodení, ako sú rozmazanie, šum, kompresia a zakrytie hornej alebo spodnej časti tváre.

Počas overovania modelov na oddelených snímkach vytvoreného datasetu dosiahli všetky modely vysokú presnosť predikcií. Avšak, pri empirickom testovaní modely preukázali čiastočnú funkcionálnu správnosť pri správnom predikovaní poškodení mimo snímok vytvoreného datasetu. Tento jav je najpravdepodobnejšie spôsobený nedostatočnou variabilitou snímok a vysokým počtom parametrov siete, čo viedlo k pretrénovaniu výsledného modelu. Ďalšími faktormi môže byť nedostatočný počet snímok v tréningovej dátovej sade, alebo nedostatočná optimalizácia hyperparametrov.

Na zlepšenie presnosti hodnotenia kvality snímok mimo použitého datasetu, by bolo vhodné rozšíriť dataset o ďalšie typy poškodení v jednotlivých kategóriách, ako napríklad rôzne typy šumu a rozmazania. Ďalším krokom pre zlepšenie hodnotenia kvality snímok by bolo rozšírenie dátovej sady o snímky rôznych rozlíšení, ktoré by mohli viesť k tomu, že sa model naučí lepšie rozpoznávať degradácie aj na snímkach mimo dátovej sady.

# Literatúra

- [1] *ISO/IEC 29794-5:2010 - Information technology - Biometric sample quality - Part 5: Face image data* International Organization for Standardization. 2010. Dostupné z: <https://www.iso.org/standard/45450.html>.
- [2] *ISO/IEC 19794-5:2011 - Information technology - Biometric data interchange formats - Part 5: Face image data* International Organization for Standardization. 2011. Dostupné z: <https://www.iso.org/standard/54193.html>.
- [3] BADILLO, S.; BANFAI, B.; BIRZELE, F.; DAVYDOV, I. I.; HUTCHINSON, L. et al. An Introduction to Machine Learning. *Clinical Pharmacology & Therapeutics*. Wiley Online Library, Apr 2020, zv. 107, č. 4, s. 871–885.
- [4] BANSAL, A.; MEHTA, K. a ARORA, S. Face Recognition Using PCA and LDA Algorithm. In: *2012 Second International Conference on Advanced Computing Communication Technologies*. 2012, s. 251–254.
- [5] BLEDSOE, W. W. *Man-machine facial recognition*. Washington, D.C.: Advanced Research Projects Agency, 1966.
- [6] BODLA, N.; SINGH, B.; CHELLAPPA, R. a DAVIS, L. S. Soft-NMS – Improving Object Detection With One Line of Code. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Oct 2017.
- [7] CABANI, A.; HAMMOUDI, K.; BENHABILES, H. a MELKEMI, M. MaskedFace-Net – A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health*, 2021, zv. 19, s. 100144. ISSN 2352-6483. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S2352648320300362>.
- [8] CABEEN, K. a GENT, P. Image Compression and the Discrete Cosine Transform. *Math 45, College of the Redwoods*, 1996.
- [9] CHEN, Y.; LI, L.; LI, W.; GUO, Q.; DU, Z. et al. Chapter 2 - Fundamentals of neural networks. In: CHEN, Y.; LI, L.; LI, W.; GUO, Q.; DU, Z. et al., ed. *AI Computing Systems*. Morgan Kaufmann, 2024, s. 17–51. ISBN 978-0-323-95399-3. Dostupné z: <https://www.sciencedirect.com/science/article/pii/B9780323953993000081>.
- [10] CHUNG, M. K. *Gaussian kernel smoothing*. 2021.
- [11] DENG, J.; GUO, J.; VERVERAS, E.; KOTSIA, I. a ZAFEIRIOU, S. RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, s. 5202–5211.

- [12] DENG, J.; GUO, J.; XUE, N. a ZAFEIRIOU, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.
- [13] DHIFLI, W. a DIALLO, A. B. Face Recognition in the Wild. *Procedia Computer Science*, 2016, zv. 96, s. 1571–1580. ISSN 1877-0509. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S1877050916320142>. Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 20th International Conference KES-2016.
- [14] DIMITROVSKI, I.; KITANOVSKI, I.; KOCEV, D. a SIMIDJIEVSKI, N. Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier BV, marec 2023, zv. 197, s. 18–35. ISSN 0924-2716. Dostupné z: <http://dx.doi.org/10.1016/j.isprsjprs.2023.01.014>.
- [15] DUMOULIN, V. a VISIN, F. *A guide to convolution arithmetic for deep learning*. 2018.
- [16] DÉNIZ, O.; BUENO, G.; SALIDO, J. a DE LA TORRE, F. Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, 2011, zv. 32, č. 12, s. 1598–1603. ISSN 0167-8655. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0167865511000122>.
- [17] ERAKLN, M. E.; DEMIR, U. a EKENEL, H. K. On Recognizing Occluded Faces in the Wild. In: IEEE. *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*. 2021, s. 1–5.
- [18] FERRARA, M.; FRANCO, A.; MAIO, D. a MALTONI, D. Face Image Conformance to ISO/ICAO Standards in Machine Readable Travel Documents. *IEEE Transactions on Information Forensics and Security*, 2012, zv. 7, č. 4, s. 1204–1213.
- [19] FREUND, Y. a SCHAPIRE, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 1997, zv. 55, č. 1, s. 119–139. ISSN 0022-0000. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S002200009791504X>.
- [20] FUNDAMENTAL RIGHTS, E. U. A. for. *Facial recognition technology: Fundamental rights considerations in the context of law enforcement*. 2019. Dostupné z: [<https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-law#publication-tab-7>]. Retrieved from [<https://fra.europa.eu/en/publication/2019/facial-recognition-technology-fundamental-rights-considerations-context-lawpublication-tab-7>].
- [21] GAN, J.; LIANG, X.; ZHAI, Y.; ZHOU, L. a WANG, B. A Real-Time Face Recognition System Based on IP Camera and SRC Algorithm. In: *Biometric Recognition - 9th Chinese Conference, CCBR 2014, Shenyang, China, November 7-9, 2014. Proceedings*. Springer, 2014, sv. 8833, s. 120–127. Lecture Notes in Computer Science. Dostupné z: [https://doi.org/10.1007/978-3-319-12484-1\\_13](https://doi.org/10.1007/978-3-319-12484-1_13).
- [22] GIRSHICK, R. *Fast R-CNN*. 2015.

- [23] GOKBERK, B.; SALAH, A.; AKARUN, L.; ETHEVE, R.; RICCIO, D. et al. 3D Face Recognition. In: April 2009, s. 263–295. ISBN 978-1-84800-291-3.
- [24] GOLDMANN, T. *face-ai-kit* <https://github.com/tgoldmann/face-ai-kit>. 2024.
- [25] GROSSI, E. a BUSCEMA, M. Introduction to artificial neural networks. *European journal of gastroenterology hepatology*, Január 2008, zv. 19, s. 1046–54.
- [26] GUO, G.; LI, S. Z. a CHAN, K. L. Support vector machines for face recognition. *Image and Vision Computing*, 2001, zv. 19, č. 9, s. 631–638. ISSN 0262-8856. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S0262885601000464>.
- [27] GUO, G.; LI, S. a CHAN, K. Face recognition by support vector machines. In: *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. 2000, s. 196–201.
- [28] HARTIGAN, J. A. a WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. JSTOR, 1979, zv. 28, č. 1, s. 100–108.
- [29] HE, K.; ZHANG, X.; REN, S. a SUN, J. *Deep Residual Learning for Image Recognition*. 2015.
- [30] HERNANDEZ ORTEGA, J.; GALBALLY, J.; FIERREZ, J.; HARAKSIM, R. a BESLAY, L. Faceqnet: Quality assessment for face recognition based on deep learning. In: IEEE. *2019 International Conference on Biometrics (ICB)*. 2019, s. 1–8.
- [31] HORÉ, A. a ZIOU, D. Image Quality Metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. 2010, s. 2366–2369.
- [32] HU, L.-H.; HUANG, M.-W.; KE, S.-W. a TSAI, C.-F. The distance function effect on k-nearest neighbor classification for medical. *Expert Systems with Applications*. Elsevier, 2011, zv. 38, č. 1, s. 804–808.
- [33] HU, X. a HUANG, B. Face Detection based on SSD and CamShift. In: *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. 2020, sv. 9, s. 2324–2328.
- [34] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *ISO/IEC 19794-5: Information technology – Biometric data interchange formats – Part 5: Face image data*. International Organization for Standardization, 2005. ISO/IEC Standard.
- [35] ITALOJS. *Facial Landmarks Recognition* <https://github.com/italojs/facial-landmarks-recognition>. 2018.
- [36] KARAHAN, S.; KILINC YILDIRUM, M.; KIRTAC, K.; RENDE, F. S.; BUTUN, G. et al. How Image Degradations Affect Deep CNN-Based Face Recognition? In: *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*. 2016, s. 1–5.
- [37] KASAR, M. M.; BHATTACHARYYA, D. a KIM, T. Face recognition using neural network: a review. *International Journal of Security and Its Applications*, 2016, zv. 10, č. 3, s. 81–100.



- [38] KIM, Y.; GWAK, H.; OH, J.; KANG, M.; KIM, J. et al. CloudNet: A LiDAR-Based Face Anti-Spoofing Model That Is Robust Against Light Variation. *IEEE Access*, 2023, zv. 11, s. 16984–16993.
- [39] KINGMA, D. a BA, J. Adam: A Method for Stochastic Optimization. In: *International Conference on Learning Representations (ICLR)*. San Diego, CA, USA: [b.n.], 2015.
- [40] LAWRENCE, S.; GILES, C.; TSOI, A. C. a BACK, A. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 1997, zv. 8, č. 1, s. 98–113.
- [41] LIU, L.; WANG, Y. a CHI, W. Image Recognition Technology Based on Machine Learning. *IEEE Access*, 2020, s. 1–1.
- [42] LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S. et al. Ssd: Single shot multibox detector. In: Springer. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. 2016, s. 21–37.
- [43] LIU, W.; WEN, Y.; RAJ, B.; SINGH, R. a WELLER, A. *SphereFace Revived: Unifying Hyperspherical Face Recognition*. 2022.
- [44] LIU, W.; WEN, Y.; YU, Z.; LI, M.; RAJ, B. et al. SphereFace: Deep Hypersphere Embedding for Face Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.
- [45] LIU, Z.; LUO, P.; WANG, X. a TANG, X. Deep Learning Face Attributes in the Wild. In: *Proceedings of International Conference on Computer Vision (ICCV)*. December 2015.
- [46] LV, H.; KO, K.; GWAK, H.; THOUMMALA, N.; KWON, H. et al. SqueezeFace: Integrative Face Recognition Methods with LiDAR Sensors. *Journal of Sensors*, 2021, zv. 2021, s. 4312245. ISSN 1687-725X. Dostupné z: <https://doi.org/10.1155/2021/4312245>.
- [47] MENG, Q.; ZHAO, S.; HUANG, Z. a ZHOU, F. MagFace: A Universal Representation for Face Recognition and Quality Assessment. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021, s. 14225–14234.
- [48] MURALEEDHARAN, V. *What is Linear Discriminant Analysis (LDA)?* <https://vivekmuraleedharan73.medium.com/what-is-linear-discriminant-analysis-lda-7e33ff59020a>. 2022. Accessed in 2023.
- [49] NWANKPA, C.; IJOMAH, W.; GACHAGAN, A. a MARSHALL, S. *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. 2018.
- [50] ODHIAMBO, M. *CelebA-HQ Resized 256x256* <https://www.kaggle.com/datasets/badasstechie/celebahq-resized-256x256/data>. 2021. Accessed: 2024.
- [51] OPENCV CONTRIBUTORS. *OpenCV modules* <https://docs.opencv.org/4.x/index.html>. 2024.

- [52] PANKAJ RAKHEJA, R. V. Image Denoising using Combination of Median Filtering and Wavelet Transform. *International Journal of Computer Applications*. New York, USA: Foundation of Computer Science (FCS), NY, USA, May 2016, zv. 141, č. 9, s. 31–35. ISSN 0975-8887. Dostupné z: <https://ijcaonline.org/archives/volume141/number9/24815-2016909803/>.
- [53] PULFER, E. *Different Approaches to Blurring Digital Images and Their Effect on Facial Detection* Computer Science and Computer Engineering Undergraduate Honors Theses. 2019. Dostupné z: <https://scholarworks.uark.edu/csceugt/66>.
- [54] REDMON, J.; DIVVALA, S.; GIRSHICK, R. a FARHADI, A. You Only Look Once: Unified, Real-Time Object Detection. In: Jún 2016, s. 779–788.
- [55] SANG, J.; LEI, Z. a LI, S. Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5. In: Jún 2009, s. 229–238. ISBN 978-3-642-01792-6.
- [56] SCHAPIRE, R. E. *Explaining AdaBoost*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. 37–52 s. ISBN 978-3-642-41136-6. Dostupné z: [https://doi.org/10.1007/978-3-642-41136-6\\_5](https://doi.org/10.1007/978-3-642-41136-6_5).
- [57] SCHROFF, F.; KALENICHENKO, D. a PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, s. 815–823.
- [58] SELAMI, A. Abd Al-salam a FADHIL, A. A Study of the Effects of Gaussian Noise on Image Features. *Kirkuk University Journal / Scientific Studies (1992-0849)*, Apríl 2016, zv. 11, s. 152 – 169.
- [59] SHAN, T.; CHEN, S.; SANDERSON, C. a LOVELL, B. Towards robust face recognition for Intelligent-CCTV based surveillance using one gallery image. In: Október 2007, s. 470–475. ISBN 978-1-4244-1696-7.
- [60] SHANG HUNG, L. An Introduction to Face Recognition Technology. *Informing Science The International Journal of an Emerging Transdiscipline*, Január 2000, zv. 3.
- [61] SHEIKH, H.; BOVIK, A. a CORMACK, L. No-reference quality assessment using natural scene statistics: JPEG2000. *Image Processing, IEEE Transactions on*, December 2005, zv. 14, s. 1918 – 1927.
- [62] SHIPITKO, O. a GRIGORYEV, A. Gaussian filtering for FPGA based image processing with High-Level Synthesis tools. In: Jún 2018.
- [63] SIMONYAN, K. a ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv 1409.1556*, September 2014.
- [64] SOLIMAN, H.; SALEH, A. a FATHY, E. Face Recognition in Mobile Devices. *International Journal of Computer Applications*, Júl 2013, zv. 73, s. 13–20.
- [65] TERVEN, J.; CORDOVA ESPARZA, D. M.; RAMIREZ PEDRAZA, A. a CHAVEZ URBIOA, E. A. *Loss Functions and Metrics in Deep Learning*. 2023.

- [66] VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Springer New York, NY, 1999. XX, 314 s. ISBN 978-0-387-98780-4. Dostupné z: <https://link.springer.com/book/10.1007/978-1-4757-3264-1#book-header>.
- [67] VIOLA, P. a JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. 2001, sv. 1, s. I–I.
- [68] WANG, L. *A survey on IQA*. 2022.
- [69] WANG, Y.-Q. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line*, 2014, zv. 4, s. 128–148. <https://doi.org/10.5201/ipol.2014.104>.
- [70] WANG, Z.; BOVIK, A. C. a LU, L. Why is image quality assessment so difficult? In: *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. 2002, sv. 4, s. IV–3313–IV–3316.
- [71] WU, H.; CAO, Y.; WEI, H. a TIAN, Z. Face Recognition Based on Haar Like and Euclidean Distance. *Journal of Physics: Conference Series*. IOP Publishing, feb 2021, zv. 1813, č. 1, s. 012036. Dostupné z: <https://dx.doi.org/10.1088/1742-6596/1813/1/012036>.
- [72] ZHANG, L.; SHEN, Y. a LI, H. VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment. *IEEE Transactions on Image Processing*, 2014, zv. 23, č. 10, s. 4270–4281.
- [73] ZHANG, L.; ZHANG, L.; MOU, X. a ZHANG, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*, 2011, zv. 20, č. 8, s. 2378–2386.
- [74] ZHU, J.; FANG, Y.; JI, P.; ABDL, M.-E. a DAI, W. RRAR: A novel reduced-reference IQA algorithm for facial images. In: *2011 18th IEEE International Conference on Image Processing*. 2011, s. 3313–3316.
- [75] ZHU, Y.; DAI, Y.; HAN, K.; WANG, J. a HU, J. An efficient bicubic interpolation implementation for real-time image processing using hybrid computing. *Journal of Real-Time Image Processing*, September 2022, zv. 19, s. 1–13.

# Príloha A

## Obsah SD

Priložené médium obsahuje nasledujúce súbory:

- `thesis-latex/` zdrojový  $\text{\LaTeX}$  kód tejto práce
- `src/` implementácia zdrojových súborov
  - `create_dataset/` zdrojové súbory pre vytvorenie datasetu
  - `resnet/` zdrojové súbory implementujúce model
  - `additional_testing/` zdrojové súbory pre testovanie modelu na iných snímkach
  - `flask_app/` zdrojové súbory výslednej aplikácie
- `data/`
  - `FIQVD/` súbory FIQVD
  - `additional_testing_data/` dodatočné testovacie súbory
- `models/` súbory natrénovaných modelov
- `thesis.pdf` pdf súbor tejto práce