

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Evidenčné číslo: FEI-104376-111119

AUTENTIFIKÁCIA EMÓCIÍ OPERÁTORA NA ZÁKLADE
VÝRAZU TVÁRE
DIPLOMOVÁ PRÁCA

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Evidenčné číslo: FEI-104376-111119

AUTENTIFIKÁCIA EMÓCIÍ OPERÁTORA NA ZÁKLADE
VÝRAZU TVÁRE
DIPLOMOVÁ PRÁCA

Študijný program:	Robotika a kybernetika
Názov študijného odboru:	kybernetika
Školiace pracovisko:	Ústav robotiky a kybernetiky
Vedúci záverečnej práce:	prof. Ing. Jarmila Pavlovičová, PhD.
Konzultant:	Ing. Michal Tölgyessy, PhD.

Bratislava 2025

Bc. Maroš Kocúr



ZADANIE DIPLOMOVEJ PRÁCE

Študent: **Bc. Maroš Kocúr**
ID študenta: 111119
Študijný program: robotika a kybernetika
Študijný odbor: kybernetika
Vedúci práce: Ing. Michal Tölgyessy, PhD.
Vedúci pracoviska: prof. Ing. František Duchoň, PhD.
Miesto vypracovania: Ústav robotiky a kybernetiky

Názov práce: **Autentifikácia emócií operátora na základe výrazu tváre**

Jazyk, v ktorom sa práca vypracuje: slovenský jazyk

Špecifikácia zadania:

Pri interakcii človeka s robotickým systémom je dôležité, aby robot vedel aj to, v akom emočnom rozpoložení sa operátor nachádza. Cieľom práce je vytvoriť modul, ktorý robotu takúto informáciu poskytne. Predpokladá sa využitie RGB kamery, ale je možnosť využiť aj RGB-D kameru.

Úlohy práce:

1. Analyzujte existujúce metódy analýzy emócií na základe výrazu tváre.
2. Naštudujte princípy tvorby biometrických modelov tváre a metódy detekcie a rozpoznávania tváre.
3. Navrhňte a implementujte systém pre identifikáciu emócií operátora na základe jeho tváre.
4. Testujte a validujte systém na simulovaných aj reálnych dátach.
6. Vytvorte ROS2 balík pre daný systém.
7. Vyhodnoťte experimenty a dosiahnuté výsledky.

Termín odovzdania diplomovej práce: 16. 05. 2025

Dátum schválenia zadania diplomovej práce: 16. 10. 2024

Zadanie diplomovej práce schválil: prof. Ing. Jarmila Pavlovičová, PhD. – garantka študijného programu

SÚHRN

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Študijný program:	Robotika a kybernetika
Autor:	Bc. Maroš Kocúr
Diplomová práca:	Autentifikácia emócií operátora na základe výrazu tváre
Vedúci záverečnej práce:	prof. Ing. Jarmila Pavlovičová, PhD.
Konzultant:	Ing. Michal Tölgyessy, PhD.
Miesto a rok predloženia práce:	Bratislava 2025

Interakcia človeka s robotom v dynamickom prostredí si čoraz viac vyžaduje pochopenie emocionálneho stavu operátora s cieľom optimalizovať komunikáciu a rozhodovacie procesy. Cieľom tejto práce je navrhnúť a implementovať modul, ktorý poskytuje robotickému systému emocionálnu spätnú väzbu a umožňuje mu zisťovať výrazy tváre operátora a odvodzovať jeho emocionálne stavy. S využitím kamery RGB s možnosťou integrácie kamery RGB-D bude systém využívať biometrické modely tváre a techniky rozpoznávania tváre na identifikáciu emócií v reálnom čase. Medzi kľúčové úlohy patrí analýza súčasných metód detekcie emócií výrazu tváre, štúdium princípov tvorby biometrických modelov tváre a implementácia robustného systému na detekciu emócií. Systém bude overený prostredníctvom testovania na simulovaných aj reálnych súboroch údajov. Okrem toho bude vyvinutý balík ROS2, ktorý zabezpečí bezproblémovú integráciu v rámci robotických systémov. Výsledky budú kriticky posúdené prostredníctvom experimentov s cieľom zabezpečiť presnosť a efektívnosť výkonu v reálnych aplikáciách.

Kľúčové slová: RGB kamera, neurónová sieť

ABSTRACT

SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

Study Programme:	Robotics and cybernetics
Author:	Bc. Maroš Kocúr
Master's thesis:	Operator Emotion Authentication Based on Facial Expression
Supervisor:	prof. Ing. Jarmila Pavlovičová, PhD.
Consultant:	Ing. Michal Tölgyessy, PhD.
Place and year of submission:	Bratislava 2025

Human-robot interaction in dynamic environments increasingly requires an understanding of the operator's emotional state to optimize communication and decision-making processes. This work aims to design and implement a module that provides emotional feedback to a robotic system, enabling it to detect the operator's facial expressions and infer emotional states. Leveraging an RGB camera, with the option to integrate an RGB-D camera, the system will employ biometric facial models and facial recognition techniques to identify emotions in real-time. Key tasks include analyzing current facial expression emotion detection methods, studying the principles of facial biometric model creation, and implementing a robust system for emotion detection. The system will be validated through testing on both simulated and real datasets. Additionally, a ROS2 package will be developed to ensure seamless integration within robotic systems. The outcomes will be critically assessed through experiments to ensure accuracy and performance efficiency in real-world applications.

Keywords: RGB camera, neural network

Pod'akovanie

I would like to express a gratitude to my thesis supervisor.

Obsah

Úvod	1
1 Úvod	2
1.1 Motivácia	3
1.2 Ciele práce	4
2 Teoretické základy	6
2.1 Emócie a ich prejav	6
2.1.1 Univerzálne emócie	6
2.1.2 Kultúrne rozdiely v prejave emócií	7
2.1.3 Výrazy tváre ako indikátory emócií	7
2.2 Analýza obrazu	8
2.2.1 Detekcia tváre	8
2.2.2 Extrakcia príznakov	9
2.2.3 Klasifikácia	9
2.3 Biometria	10
2.3.1 Princípy biometrických systémov	10
2.3.2 Identifikácia vs. verifikácia	11
2.3.3 Výhody a výzvy biometrie tváre	11
3 Existujúce metódy analýzy emócií	12
3.1 Ručne značenie	12
3.2 Automatická analýza emócií	12
3.2.1 Konvolučné neurónové siete	12
3.2.2 Typy vhodných neurónových sietí	13
3.2.3 Príklady použitia počítačového videnia	13
4 Návrh riešenia	15
4.1 Porovnanie vybraných metód	15
4.2 Architektúra systému ResEmoteNet	16
4.2.1 Hlavné komponenty architektúry	16
4.2.2 Prínosy architektúry ResEmoteNet	17
4.2.3 Experimentálne výsledky	17
4.2.4 Schéma architektúry systému	17
4.3 Výber dát	19

4.3.1	Dôvody výberu RAF-DB	19
4.3.2	Porovnanie s inými datasetmi	19
4.3.3	Prínos pre prácu	19
4.3.4	Príprava dát	20
4.4	Extrakcia príznakov	20
4.4.1	Metódy extrakcie príznakov	20
4.4.2	Porovnanie metód	21
4.4.3	Integrácia príznakov v ResEmoteNet	21
4.4.4	Experimentálne výsledky	22
4.5	Klasifikácia	22
4.5.1	Metódy klasifikácie	22
4.5.2	Porovnanie metód klasifikácie	23
4.5.3	Implementácia klasifikátora v systéme ResEmoteNet	23
4.5.4	Výsledky klasifikácie na datasete RAF-DB	23
4.6	Výber hyperparametrov	24
4.6.1	Kľúčové hyperparametre a ich úloha	24
4.6.2	Porovnanie vplyvu hyperparametrov	25
4.6.3	Optimalizačné stratégie	25
4.6.4	Augmentačné parametre	25
4.6.5	Záver	25
4.7	Integrácia do robotického pracoviska COCOHRIP	26
5	Implementácia riešenia	27
5.1	Vývojové prostredie a infraštruktúra	27
5.2	Trénovanie modelu ResEmoteNet	27
5.2.1	Organizácia vývojového prostredia	27
5.2.2	Načítavanie a príprava dát	27
5.2.3	Augmentácia dát	27
5.2.4	Architektúra modelu	27
5.2.5	Trénovací proces	28
5.2.6	Ukladanie checkpointov	28
5.2.7	Vizualizácia výsledkov	28
5.2.8	Optimalizačné výzvy	28
5.3	Integrácia do ROS2 ekosystému	29
5.4	Webová vizualizácia pomocou Flask	29

5.5	Validácia na robotickom pracovisku COCOHRIP	29
6	Exprerimenty a vyhodnotenie	30
6.1	Dátová sada	30
6.2	Metriky	30
6.3	Výsledky	31
6.4	Analýza výsledkov	33
7	Záver	34
7.1	Zhodnotenie práce	34
7.2	Obmedzenia práce	34
7.3	Budúce smerovanie	34
8	Doplnujece poznámky	35
9	Plán práce na ďalší semester	36
	Záver	37
	Zoznam použitej literatúry	38
	Prílohy	40

Zoznam obrázkov a tabuliek

Obrázok 1.1	Schéma systému na rozpoznávanie emócií operátora pomocou RGB kamery.	5
Obrázok 2.1	Detekcia tváre pomocou Haar Cascade.	9
Obrázok 2.2	Detekcia tváre pomocou SSD modelu.	9
Obrázok 4.1	porovnanie rôznych metód na rozpoznávanie emócií [11].	16
Obrázok 4.2	Architektúra systému na rozpoznávanie emócií operátora pomocou RGB kamery.	18
Obrázok 4.3	Príklad vizualizácie feature máp v rôznych vrstvách ResEmoteNet .	21
Obrázok 4.4	Architektúra COCOHRIP s vyznačením pozície emocionálneho modulu	26
Obrázok 6.1	Confusion matrix.	30
Obrázok 6.2	Trénovacie grafy.	31
Obrázok 6.3	Examples of different facial expressions.	32
Tabuľka 4.1	Porovnanie vybraných datasetov na rozpoznávanie emócií	19
Tabuľka 4.2	Porovnanie metód extrakcie príznakov	21
Tabuľka 4.3	Porovnanie metód klasifikácie emócií	23
Tabuľka 4.4	Vplyv hyperparametrov na presnosť modelu	25

Zoznam skratiek

CNN	Convolutional neural network
DCNN	Deep convolutional neural network
LSTM	Long short-term memory
RNN	Recurrent neural network

Zoznam algoritmov

Zoznam výpisov

1	Spustacie skripty systému	29
---	-------------------------------------	----

Úvod

Ukazka upraveného šablony FEIStyle.cls (<https://github.com/Kyslik/FEIStyle>) s použitím Times New Roman fontu. Nastavenie fontov som prebral z oficiálneho IEEE šablony, vložil som ho do FEIStyle.cls na riadky 228-230.

Tu bude krásny úvod s diakritikou atď.

A možno aj viac riadkový úvod.

1 Úvod

S rozvojom umelej inteligencie a strojového učenia sa otvárajú nové možnosti pre interakciu medzi človekom a strojom. Jednou z najdôležitejších oblastí výskumu je rozpoznávanie emócií na základe výrazu tváre, ktoré umožňuje strojom porozumieť emocionálnemu stavu používateľ'a. V kontexte robotických systémov je dôležité, aby roboty boli schopné rozoznať emócie človeka, čo môže zlepšiť komunikáciu, kooperáciu a bezpečnosť pri spoločnej práci [1].

Psychológ Paul Ekman pomenoval šesť základných emócií - šťastie, smútok, hnev, strach, prekvapenie a znechutenie - ktoré sú často základom pre zavedené kategórie emócií pri rozpoznávaní tváre. Výber týchto emócií bol založený na ich univerzálnom rozpoznávaní a pozorovaní naprieč kultúrami, čo ich predurčuje na použitie v systémoch rozpoznávania tváre.

Podľa Ekmanovho výskumu sa tieto pocity odrážajú v konkrétnych výrazoch tváre, ktoré dokážu automaticky rozpoznať ľudia zo všetkých kultúrnych prostredí. Jeho práca vytvorila fundamentálny základ pre strojové učenie a psychológiu, najmä pri vytváraní systémov, ktoré dokážu dešifrovať výrazy tváre na určenie emocionálneho stavu jednotlivca.

V záujme konzistentnosti a presnosti v aplikáciách, ako je zdravotníctvo, robotika a služby zákazníkom, je možné do systémov rozpoznávania tváre zahrnúť konzistentnú metódu analýzy emócií. Pochopenie úrovne spokojnosti alebo podráždenia používateľ'a môže napríklad pomôcť upraviť reakcie systému a zlepšiť výsledky interakcií [2].

Emócie zohrávajú dôležitú úlohu v procese rozhodovania, riadenia a interakcie. Schopnosť robotického systému porozumieť emocionálnemu stavu používateľ'a umožňuje jeho prispôbenie konkrétnym podmienkam a potrebám operátora. Napríklad v priemysle môžu robotické systémy identifikovať stres alebo únavu operátora, čím prispievajú k zvýšeniu bezpečnosti a efektivity. Okrem toho, v oblasti zdravotnej starostlivosti môže rozpoznávanie emócií pomôcť monitorovať psychický stav pacientov a prispieť k ich lepšej starostlivosti [3].

Rozpoznávanie emócií je možné dosiahnuť rôznymi metódami, ktoré zahŕňajú spracovanie obrazu, analýzu textu, reč a gestá. Výraz tváre je však najvýznamnejším a najpresnejším indikátorom emócií, pretože vyjadruje okamžitý emocionálny stav človeka. Emócie, ako sú šťastie, smútok, hnev alebo prekvapenie, sú viditeľné prostredníctvom zmien vo svaloch tváre, ktoré sú merateľné a analyzovateľné pomocou technológií strojového učenia, najmä pomocou hlbokých neurónových sietí (Convolutional neural network (CNN)) [1].

Súčasná metóda na rozpoznávanie emócií zahŕňa viacero prístupov. Tradičné prístupy, ako napríklad metódy založené na geometrických črtách a textúrach, boli doplnené modernými metódami založenými na hlbokom učení, ktoré dosahujú vysokú presnosť. Neurónové siete sú

schopné automaticky extrahovať črty tváre bez potreby manuálneho zásahu, čo výrazne zvyšuje efektivitu systému. Tieto pokročilé modely dosahujú vysokú mieru úspešnosti v rôznych aplikáciách, ako sú zdravotná starostlivosť, priemyselná automatizácia alebo monitorovanie únavy vodičov [3] [4].

1.1 Motivácia

V súčasnej dobe, keď technológie prenikajú do všetkých aspektov ľudského života, je nevyhnutné, aby robotické systémy disponovali nielen kognitívnymi schopnosťami, ale aj emocionálnou inteligenciou. Táto schopnosť umožňuje robotom identifikovať, hodnotiť a reagovať na emócie ľudí, čím sa zvyšuje efektivita a prirodzenosť interakcie medzi človekom a strojom. Emocionálna inteligencia zahŕňa schopnosť rozpoznávať vlastné emócie, emócie iných, a adekvátne na ne reagovať.

V oblasti interakcie človeka s robotom (Human-Robot Interaction, HRI) je rozpoznávanie emócií kľúčové pre vytvorenie intuitívnej a efektívnej spolupráce. Roboty schopné interpretovať emocionálny stav operátora môžu prispôbiť svoje správanie aktuálnej situácii, čo je obzvlášť dôležité v oblastiach, kde je potrebná vysoká miera spolupráce a dôvery, ako napríklad v zdravotníctve, vzdelávaní či v priemyselnej výrobe. Napríklad v zdravotníctve môžu roboty asistovať pacientom s emocionálnou podporou, monitorovať ich psychický stav a poskytovať adekvátne reakcie na základe rozpoznaných emócií [5].

Aktuálne trendy v oblasti robotiky smerujú k vývoju systémov, ktoré dokážu nielen vykonávať predprogramované úlohy, ale aj adaptovať sa na dynamické prostredie a emocionálne potreby používateľov. Výskum v oblasti afektívnej robotiky sa zameriava na integráciu emocionálnych modelov do robotických systémov, čo umožňuje lepšie porozumenie a predikciu ľudského správania. Napriek pokroku v tejto oblasti existujú výzvy spojené s presnosťou rozpoznávania emócií, interpretáciou komplexných emocionálnych stavov a zabezpečením etického využitia týchto technológií [5].

Implementácia systému na identifikáciu emócií operátora na základe výrazu tváre predstavuje významný krok k zlepšeniu interakcie medzi človekom a robotom. Takýto systém môže prispieť k vyššej efektivite, bezpečnosti a spokojnosti používateľov pri práci s robotickými systémami. Napríklad v priemyselných aplikáciách môže robot upraviť svoj pracovný rytmus alebo poskytnúť upozornenie v prípade, že deteguje stres alebo únavu operátora, čím sa predchádza možným chybám alebo nehodám.

Celkovo je motivácia pre vývoj takýchto systémov založená na snahe o vytvorenie robotických asistentov, ktorí sú schopní nielen technicky podporovať človeka, ale aj empaticky reagovať na jeho emocionálne potreby, čím sa zvyšuje kvalita a efektivita ich vzájomnej spo-

lupráce. Motiváciou pre rozpoznávanie emócií tváre je jeho potenciál zlepšiť interakciu medzi človekom a počítačom, zlepšiť monitorovanie duševného zdravia a vytvoriť adaptívne systémy pre rôzne oblasti, ako je vzdelávanie, marketing a robotika [5].

1.2 Ciele práce

Cieľom tejto práce je navrhnúť, implementovať a otestovať systém, ktorý bude schopný identifikovať emocionálny stav človeka na základe výrazu jeho tváre v reálnom čase. Takýto systém má potenciál výrazne zlepšiť interakciu človeka s robotickým systémom, najmä v prostredí, kde je dôležitá adaptácia robota na aktuálny psychický stav operátora.

Hlavný dôraz je kladený na vytvorenie robustného a presného modulu na rozpoznávanie emócií, ktorý bude možné integrovať do existujúcich robotických platforiem v prostredí ROS2 (Robot Operating System 2). V práci sa uvažuje s využitím RGB kamery, avšak je otvorený priestor aj pre rozšírenie o RGB-D kamery (napr. RealSense alebo ZED), ktoré umožňujú snímanie hĺbky a tým zvyšujú presnosť 3D analýzy tváre.

Práca sa zameriava na splnenie nasledujúcich čiastkových cieľov:

Analýza existujúcich metód rozpoznávania emócií na základe výrazu tváre – preskúmanie prístupov založených na tradičnom strojovom učení (SVM, HOG + LBP), ako aj moderných hlbokých neurónových sietí (CNN, ResNet, SE-ResNet, atď.).

Štúdium biometrických modelov tváre a techník detekcie tváre – výskum metód ako Haar Cascade, MTCNN, Dlib a OpenCV moduly na detekciu a orezanie tváre.

Návrh architektúry systému na rozpoznávanie emócií – vybudovanie modelu (napr. ResEmoteNet) využívajúceho konvolučné siete a reziduálne bloky, schopného klasifikovať emócie ako radosť, smútok, hnev, prekvapenie, odpor a strach.

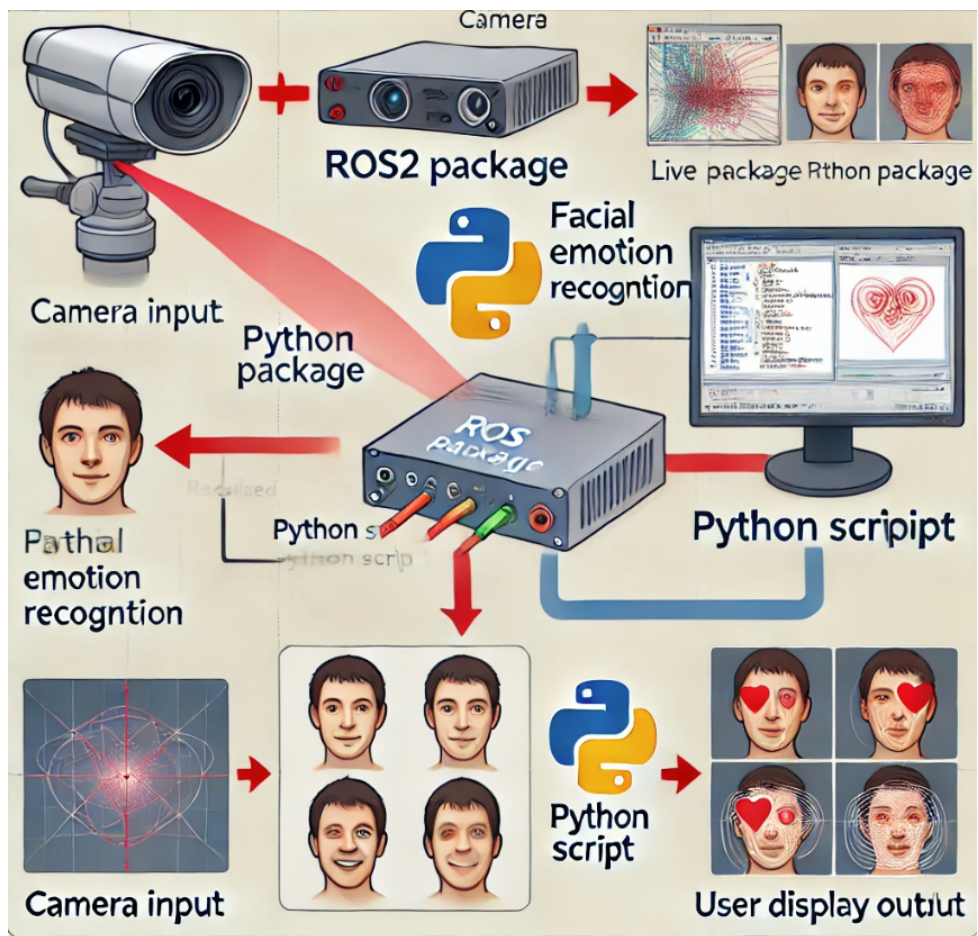
Tréning a testovanie systému na open-source databázach – použitie datasetov ako FER2013, RAF-DB, AffectNet a KDEF na overenie presnosti a generalizácie modelu.

Vytvorenie ROS2 balíka – implementácia systému do samostatného ROS2 balíka, ktorý poskytne výstup emocionálneho stavu ako ROS2 téma (napr. emotion_topic).

Validácia a experimentálne vyhodnotenie systému – testovanie modelu na reálnych aj simulovaných dátach (napr. v Gazebo alebo RViz), s dôrazom na presnosť, rýchlosť spracovania a robustnosť voči zmenám osvetlenia či polohy tváre.

Cieľom je vytvoriť systém, ktorý bude nielen presný a spoľahlivý, ale aj dostatočne efektívny na to, aby mohol byť nasadený v reálnom čase na vstavaných výpočtových zariadeniach ako NVIDIA Jetson alebo Raspberry Pi 4 a COCOHRIP. Systém by mal byť zároveň jednoducho škálovateľný a rozšíriteľný o ďalšie vstupy (napr. hlas, srdcovú frekvenciu alebo EDA senzory), čo otvorí cestu k multimodálnemu rozpoznávaniu emócií. Cieľom práce je vytvoriť

system, ktorý bude schopný rozpoznat' emócie v reálnom čase.



Obr. 1.1: Schéma systému na rozpoznávanie emócií operátora pomocou RGB kamery.

2 Teoretické základy

2.1 Emócie a ich prejav

Emócie predstavujú základný prvok ľudskej psychiky a zohrávajú kľúčovú úlohu pri vnímaní, rozhodovaní a správaní človeka. Ide o komplexné psychologické stavy, ktoré sú výsledkom interakcie medzi subjektívnymi prežitkami, fyziologickými reakciami a behaviorálnymi prejavmi. Emócie sú neoddeliteľnou súčasťou medzilidskej komunikácie a výrazne ovplyvňujú spôsob, akým človek interpretuje a reaguje na svoje okolie [5].

V oblasti rozpoznávania emócií je dôležité chápať, že emocionálne stavy sa môžu prejavovať prostredníctvom viacerých modalít – výrazu tváre, reči, gest, telesného pohybu a fyziologických parametrov. V tejto práci sa zameriavame predovšetkým na výraz tváre, keďže ide o najvýraznejší a najintuitívnejší indikátor emocionálneho stavu, ktorý je zároveň vhodný na analýzu pomocou techník počítačového videnia.

Podľa výskumu psychológa Paula Ekmana existuje šesť základných emócií, ktoré sú univerzálne rozpoznateľné na základe výrazu tváre naprieč kultúrami: radosť, smútok, hnev, strach, prekvapenie a odpor. Tieto emócie majú charakteristické črty, ktoré sú podmienené aktivitou konkrétnych svalových skupín v tvári. Napríklad radosť sa často prejavuje zdvihnutím kútikov úst a vytvorením vrások okolo očí, zatiaľ čo hnev býva charakterizovaný znížením obočia a stiahnutými perami.

Význam výrazu tváre ako neverbálneho komunikačného kanála podčiarkuje aj výskum, ktorý ukazuje, že až 55 % emocionálnych informácií v interpersonálnej komunikácii je sprostredkovaných cez tvár. Tento údaj zdôrazňuje dôležitosť správnej analýzy mimiky pri snahe strojovo identifikovať emócie.

Z praktického hľadiska má analýza výrazu tváre široké uplatnenie. V oblasti zdravotnej starostlivosti môže pomáhať monitorovať psychický stav pacientov, v automotive sektore identifikovať únavu alebo stres vodiča, a v robotike zvyšovať adaptabilitu robotických systémov pri práci s ľuďmi.

Z výskumného pohľadu je preto nevyhnutné dôkladne porozumieť tomu, ako sa emócie prejavujú vo výraze tváre, a akými algoritmickými prístupmi je možné tieto prejavy detegovať a interpretovať. To tvorí východisko pre praktickú časť práce, kde sa tieto teoretické poznatky pretavia do konkrétneho technického riešenia.

2.1.1 Univerzálne emócie

Jednou z najvýznamnejších teórií o emóciách je teória univerzálnych emócií, ktorú vyvinul psychológ Paul Ekman. Podľa tejto teórie existuje šesť základných emócií, ktoré sú univerzálne

rozpoznateľné na základe výrazu tváre: radosť, smútok, hnev, prekvapenie, strach a odpor. Tieto emócie sú nezávislé od kultúrnych vplyvov a prejavujú sa podobným spôsobom naprieč rôznymi kultúrami a etnickými skupinami [3].

2.1.2 Kultúrne rozdiely v prejave emócií

Aj napriek existencii univerzálnych emócií existujú značné kultúrne rozdiely v tom, ako sú emócie prejavované a interpretované. Kultúrne normy a spoločenské očakávania môžu výrazne ovplyvniť intenzitu, frekvenciu a spôsob vyjadrenia emócií.

V tzv. individualistických kultúrach (napr. západná Európa alebo USA) je emocionálny prejav väčšinou priamy a otvorený. Naopak, v kolektivistických kultúrach (napr. východná Ázia) býva prejav emócií častejšie potláčaný alebo upravený v záujme zachovania harmónie v skupine.

Tieto kultúrne odlišnosti predstavujú výzvu pre univerzálne systémy rozpoznávania emócií, pretože rovnaký výraz tváre môže byť interpretovaný odlišne v závislosti od kultúrneho kontextu. Z tohto dôvodu je v niektorých prípadoch výhodné využiť aj regionálne adaptované modely, alebo multimodálne systémy, ktoré pracujú s doplnujúcimi vstupmi (reč, gestá, fyziologické signály).

2.1.3 Výrazy tváre ako indikátory emócií

Výraz tváre predstavuje najdôležitejší vizuálny prejav emocionálneho stavu človeka. Ide o dynamický proces, pri ktorom sa aktivujú rôzne svaly tváre (tzv. action units) a vytvárajú špecifické konfigurácie typické pre jednotlivé emócie. Napríklad:

Action Unit 12 (zdvihnutie kútikov úst) je typické pre radosť,

Action Unit 4 (stiahnutie obočia) pre hnev,

Action Unit 1 + 2 (zdvihnutie vnútorných obočí) pre smútok.

Tieto jednotky boli formalizované v rámci systému FACS (Facial Action Coding System), ktorý vyvinul Ekman spolu s Friesenom. FACS poskytuje štandardizovaný spôsob, ako kvantifikovať a analyzovať pohyby svalov tváre, čím sa stal základom pre mnohé moderné algoritmy rozpoznávania emócií.

Z výskumov vyplýva, že až 55 – 60 % emocionálnych informácií je komunikovaných prostredníctvom výrazu tváre, čo len podčiarkuje jeho význam v oblasti afektívneho počítania. Z tohto dôvodu sú technológie počítačového videnia a strojového učenia primárne zamerané práve na vizuálnu modalitu tváre, ktorá poskytuje spoľahlivý a relatívne ľahko spracovateľný vstup pre automatizované systémy [6].

2.2 Analýza obrazu

Analýza obrazu je kľúčovým prvkom systému rozpoznávania emócií, pretože umožňuje spracovať vizuálny vstup a získať z neho relevantné črty tváre, ktoré sú následne použité na klasifikáciu emocionálneho stavu. Tento proces sa zvyčajne skladá z troch základných krokov:

- **Detekcia tváre** – určenie polohy tváre v obraze,
- **Extrakcia príznakov** – získanie vizuálnych črt z detegovanej tváre,
- **Klasifikácia** – priradenie emócie na základe získaných črt.

Tieto kroky je možné realizovať pomocou klasických metód počítačového videnia (napr. OpenCV), ako aj s využitím moderných hlbokých neurónových sietí (napr. CNN), ktoré často spájajú všetky tri fázy do jedného end-to-end modelu.

2.2.1 Detekcia tváre

Detekcia tváre je prvým krokom pri analýze obrazu. Ide o proces lokalizácie oblasti tváre v snímke, ktorý slúži ako základ pre ďalšie spracovanie. Existujú dve základné kategórie prístupov: klasické (ručne definované príznaky) a moderné (hlboké učenie).

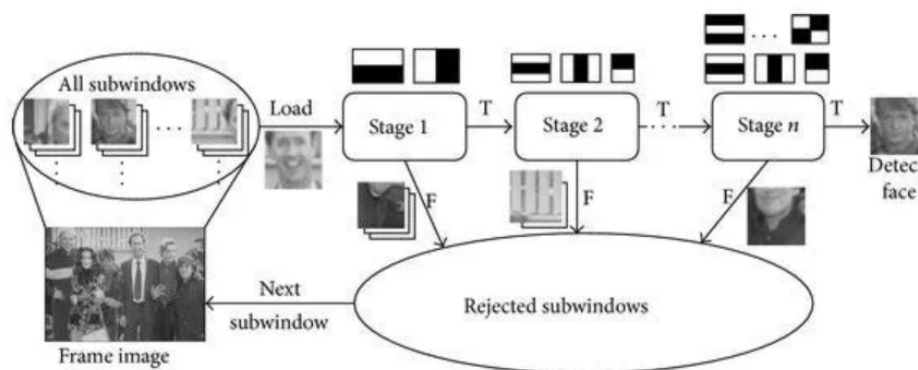
Viola-Jones (Haar Cascade) Algoritmus Viola-Jones využíva tzv. *Haar-like* príznaky a Ada-boost klasifikátor trénovaný na veľkom množstve tvárí. Výhodou je nízka výpočtová náročnosť, vďaka čomu je vhodný aj pre zariadenia s obmedzeným výkonom. Slabinou je citlivosť na uhol natočenia tváre a osvetlenie.

SSD Face Detection Metóda *Single Shot MultiBox Detector (SSD)* patrí medzi hlboké konvolučné modely schopné detegovať tváre v reálnom čase. Vstupný obraz sa normalizuje a spracúva sieťou, ktorá výstupom poskytuje ohraničujúce boxy a skóre istoty (*confidence score*). Výhodou je schopnosť robustnej detekcie aj pri zakrytí časti tváre alebo v zhoršených svetelných podmienkach.

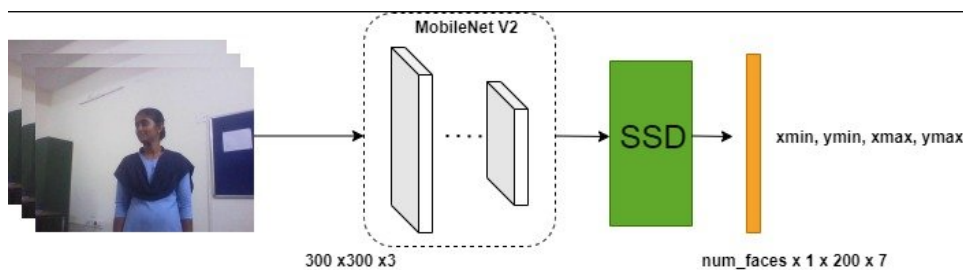
Iné pokročilé metódy Ďalšie populárne detektory zahŕňajú:

- **MTCNN** (Multi-task CNN): detekcia spolu s lokalizáciou kľúčových bodov tváre,
- **YOLO-Face**: adaptácia YOLO modelu pre vysokorýchlostnú detekciu tvárí,
- **RetinaFace**: detektor schopný presne odhadnúť aj pózu a tvar tváre.

Moderné metódy sa bežne trénujú na datasetoch ako WIDER-Face alebo FDDB a dosahujú vysokú presnosť i rýchlosť.



Obr. 2.1: Detekcia tváre pomocou Haar Cascade.



Obr. 2.2: Detekcia tváre pomocou SSD modelu.

2.2.2 Extrakcia príznačkov

Po úspešnej detekcii tváre nasleduje fáza extrakcie príznačkov (*feature extraction*), kde sa analyzujú črty ako oči, ústa, obočie, nos a ich relatívne pozície. Existujú dva hlavné prístupy:

- **Geometrické príznačky:** meranie vzdialeností a uhlov medzi bodmi (napr. vzdialenosť medzi očami),
- **Textúrne príznačky:** využívajú sa filtre ako Gabor, LBP (Local Binary Patterns) alebo HOG (Histogram of Oriented Gradients).

Moderné prístupy využívajú **konvolučné neurónové siete (CNN)**, ktoré automaticky extrahujú hierarchické príznačky z obrazových dát bez potreby manuálneho návrhu črt. Tento spôsob je robustnejší voči zmenám osvetlenia, výrazu a šumu.

2.2.3 Klasifikácia

Klasifikácia je záverečným krokom v analýze obrazu, kde sa príznačky spracované z predchádzajúcej fázy priradujú k určitej kategórii emócií. V tradičných systémoch sa používali metódy ako:

- *Support Vector Machines (SVM)*,

- *Random Forest*,
- *K-Nearest Neighbors (k-NN)*.

V súčasnosti sa najčastejšie používajú hlboké neurónové siete, predovšetkým CNN, prípadne ich kombinácie s Recurrent neural network (RNN) alebo LSTM pre spracovanie časových sekvencií vo videách.

Pre potreby tejto práce bude využitý model **ResEmoteNet**, ktorý kombinuje konvolučné vrstvy, SE bloky a reziduálne bloky, čím dosahuje vysokú presnosť pri klasifikácii emócií na viacerých verejných datasetoch.

2.3 Biometria

Biometria je vedný odbor zaoberajúci sa rozpoznávaním a identifikáciou osôb na základe ich jedinečných fyziologických alebo behaviorálnych charakteristík. V kontexte tejto práce sa biometria sústreďuje predovšetkým na **biometriu tváre**, ktorá využíva špecifické črty ľudskej tváre na autentifikáciu, verifikáciu alebo identifikáciu jednotlivcov.

Rozpoznávanie tváre patrí medzi najpoužívanejšie a najprirodzenejšie biometrické techniky. Na rozdiel od iných biometrických metód (napr. odtlačky prstov alebo dúhovka), tvár je voľne dostupná a možno ju snímať neinvazívne aj bez vedomia pozorovaného subjektu, čo otvára možnosti pre pasívne monitorovanie, ale zároveň vyžaduje dôsledné riešenie otázok ochrany súkromia[7] [3].

2.3.1 Princípy biometrických systémov

Každý biometrický systém pozostáva z nasledujúcich hlavných komponentov:

- **Zachytávacie zariadenie** – typicky kamera (RGB alebo RGB-D), ktorá sníma obraz tváre.
- **Predspracovanie** – normalizácia osvetlenia, vyrovnanie orientácie a orezanie oblasti tváre.
- **Extrakcia príznakov** – získanie reprezentatívnych črt pomocou geometrických, textúrnych alebo hlbokých metód.
- **Porovnanie** – porovnanie aktuálne extrahovaných črt s referenčnými dátami (napr. v databáze).
- **Rozhodovanie** – určenie, či záznam zodpovedá známej osobe (verifikácia) alebo ktorá osoba to je (identifikácia).

Úspešnosť biometrického systému závisí od kvality vstupných dát, výberu algoritmu, robustnosti extrakcie príznakov a schopnosti pracovať s variabilitami, ako sú zmeny výrazu tváre, osvetlenia, uhla alebo čiastočného zakrytia.

2.3.2 Identifikácia vs. verifikácia

V biometrických systémoch rozlišujeme dva hlavné režimy práce:

- **Verifikácia** – odpovedá na otázku: „Je táto osoba tým, za koho sa vydáva?“ Ide o porovnanie 1:1 medzi aktuálnou tvárou a uloženou šablónou konkrétného používateľa.
- **Identifikácia** – odpovedá na otázku: „Kto je táto osoba?“ Porovnáva aktuálne údaje s celou databázou záznamov (porovnanie 1:N).

V prípade rozpoznávania emócií z výrazu tváre je typickým scenárom práve **identifikácia emocionálneho stavu**, nie identity samotnej osoby. Avšak v systémoch, kde je potrebné súčasne rozoznávať osobu a jej emócie (napr. personalizované robotické asistenty), môže byť výhodné obidve úlohy kombinovať [7].

2.3.3 Výhody a výzvy biometrie tváre

Rozpoznávanie tváre má oproti iným biometrickým metódam viacero výhod:

- *neinvazívnosť a bezkontaktnosť*,
- *vhodnosť pre sledovanie v reálnom čase*,
- *vysoká akceptovateľnosť u používateľov*,
- *možnosť kombinácie s inými modalitami (hlas, gesta, emócie)*.

Medzi hlavné výzvy patrí:

- variabilita výrazu tváre (napr. úsmev vs. hnev),
- zmeny spôsobené vekom, svetelnými podmienkami alebo pohybom,
- zakrytie tváre (napr. rúška, okuliare),
- etické a legislatívne otázky súvisiace s ochranou súkromia a spracovaním biometrických údajov.

Tieto faktory musia byť zohľadnené najmä v aplikáciách v reálnom svete, kde je požadovaná vysoká robustnosť a spoľahlivosť systému.

3 Existujúce metódy analýzy emócií

V oblasti rozpoznávania emócií na základe výrazu tváre existuje mnoho prístupov, ktoré môžeme rozdeliť na manuálne a automatizované metódy. Kým tradičné manuálne prístupy spočívajú v ručnom označovaní výrazov tváre, moderné metódy využívajú automatické algoritmy, často založené na neurónových sieťach (NN).

3.1 Ručne značenie

Ručne značenie (manuálna anotácia) spočíva v označovaní kľúčových bodov na tvári a následnom priradení výrazov tváre k určitým emočným kategóriám. Tento proces je časovo náročný a vyžaduje expertov na interpretáciu dát. Avšak, ručné značenie je stále dôležité pre tvorbu datasetov, ktoré sú nevyhnutné na tréning automatických systémov. Dôležité datasetové projekty, ako sú Cohn-Kanade alebo AffectNet, sa opierajú o ručné značenie výrazov tváre. Manuálna anotácia má významnú úlohu v počiatočných fázach výskumu, ale pre aplikácie, ktoré vyžadujú veľké množstvo dát, je neefektívna [5].

Pre kvalitný výstup modelu je nevyhnutné disponovať **presnými a konzistentnými anotáciami**. Nekonzistentné alebo chybné značkovanie môže výrazne ovplyvniť výslednú presnosť modelu, preto je vhodné, aby anotáciu vykonávali vyškolení ľudia a prípadne sa zabezpečila viacnásobná anotácia (napr. metóda majority voting).

3.2 Automatická analýza emócií

Automatická analýza emócií predstavuje pokročilý spôsob interpretácie emocionálneho stavu človeka bez potreby manuálneho zásahu. Využíva najnovšie metódy počítačového videoní a strojového učenia, ktoré umožňujú strojom rozpoznať emócie na základe výrazu tváre v reálnom čase. Cieľom takýchto systémov je identifikovať jemné zmeny v mimike a priradiť ich k zodpovedajúcej kategórii emócie.

Moderné prístupy sa vo veľkej miere spoliehajú na architektúry hlbokého učenia, predovšetkým na CNN, ktoré automaticky extrahujú relevantné črty z obrazu tváre. Kombináciou týchto sietí s rekurentnými architektúrami, ako sú RNN a Long short-term memory (LSTM), je možné efektívne analyzovať aj dynamiku výrazu tváre v čase, čím sa výrazne zvyšuje presnosť rozpoznávania pre videozáznamy a reálne situácie [5] [8].

3.2.1 Konvolučné neurónové siete

Konvolučné neurónové siete (CNN) sú inšpirované biologickými procesmi vo vizuálnom kortexe a patria medzi najefektívnejšie metódy spracovania obrazových dát. V kontexte rozpoznávania emócií sú CNN schopné automaticky extrahovať príznaky výrazu tváre bez potreby

ručnej definície, čím zjednodušujú a zrýchľujú proces spracovania. Ich viacvrstvová architektúra umožňuje postupnú extrakciu od základných črt (napr. hrany a krivky) až po komplexnejšie tvárové štruktúry ako sú oči, obočie, ústa či líce svaly.

Zásadnou výhodou CNN je ich robustnosť voči variabilite v osvetlení, uhle pohľadu a individuálnych rozdieloch medzi ľuďmi. Tieto vlastnosti z nich robia ideálnu voľbu pre reálne nasadenie do systémov rozpoznávania emócií [9].

3.2.2 Typy vhodných neurónových sietí

Pre aplikácie rozpoznávania emócií sa najčastejšie využívajú nasledovné typy hlbokých neurónových sietí:

- **CNN:** Používané na extrakciu *priestorových* črt tváre z jednotlivých obrázkov. Sú schopné zachytiť štruktúry tváre aj pri rôznych výrazoch a uhle natočenia.
- **RNN a LSTM:** Rekurentné siete sú vhodné pre spracovanie *časových sekvencií*, ako sú videozáznamy mimiky. Ich schopnosť uchovávať predchádzajúce stavy umožňuje sledovať zmeny vo výraze v čase, čo je kľúčové pri rozpoznávaní prechodových emócií alebo mikroexpresí [10].
- **Deep convolutional neural network (DCNN):** Hlboké konvolučné siete predstavujú pokročilejší variant CNN, ktorý zahŕňa väčší počet vrstiev a často aj reziduálne alebo SE bloky. Používajú sa v systémoch s dôrazom na vysokú presnosť a zložitejšiu klasifikáciu emócií v náročných podmienkach.

3.2.3 Príklady použitia počítačového videnia

Automatické rozpoznávanie emócií je úzko prepojené s oblasťou počítačového videnia. Táto disciplína sa zaoberá extrakciou a analýzou vizuálnych informácií z obrazových vstupov. CNN modely sa často kombinujú s tradičnými technikami extrakcie črt, ako sú **HOG (Histogram of Oriented Gradients)** alebo **SIFT (Scale-Invariant Feature Transform)**, ktoré pomáhajú zlepšiť robustnosť systému, najmä v prípade rušivých podmienok.

Tieto systémy sú navrhnuté tak, aby dokázali identifikovať a klasifikovať emocionálne výrazy aj v podmienkach ako sú:

- nehomogénne alebo slabé osvetlenie,
- čiastočné zakrytie tváre (napr. rúškou, rukou, okuliarmi),
- rôzne etnické alebo vekové skupiny,
- nečakané výrazy (kombinácie viacerých emócií).

Dôkazom úspešnosti týchto prístupov je ich nasadenie v reálnych aplikáciách, napríklad v zákazníckych centrách, automobilovom priemysle (detekcia únavy vodiča), v zdravotníctve (monitorovanie pacientov) či v robotike (adaptívne roboty s afektívnou spätnou väzbou) [Huang2023].

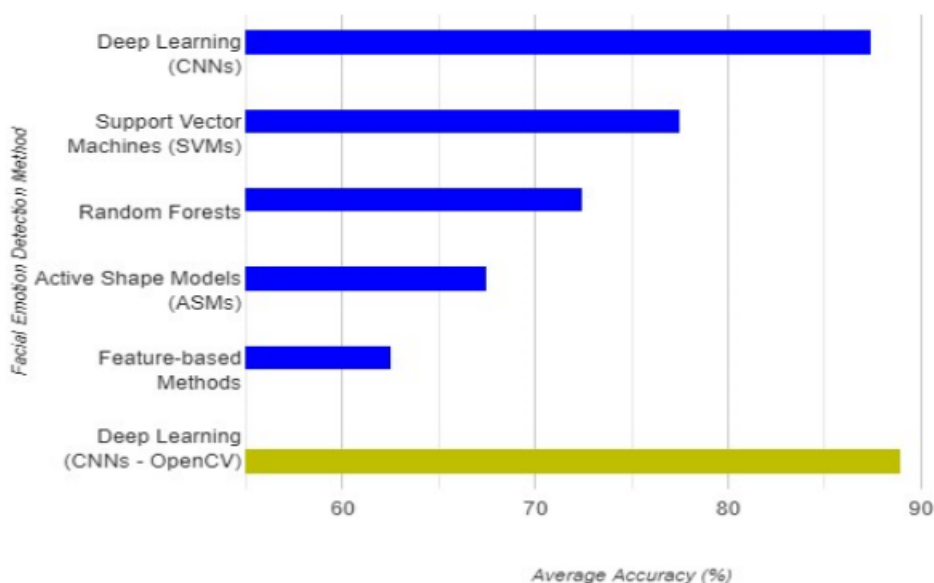
4 Návrh riešenia

4.1 Porovnanie vybraných metód

V dokumente sa ako hlavné metódy použili kombinácia OpenCV a konvolučných neurónových sietí (CNN), pričom tieto technológie boli zvolené kvôli schopnosti riešiť výzvy spojené so zmenami osvetlenia, rôznymi pózami a potrebou rýchleho spracovania v reálnom čase. OpenCV slúžil na efektívne predspracovanie dát a detekciu tvárí pomocou Haar cascade classifier, zatiaľ čo CNN umožnili extrakciu komplexných črt a klasifikáciu emócií. Ako základ bola použitá predtrénovaná architektúra VGGFace, ktorá výrazne zlepšila presnosť systému a skrátila čas potrebný na tréning. Model bol trénovaný na datasete FER2013, ktorý obsahoval obrázky kategorizované podľa siedmich základných emócií (šťastie, smútok, hnev, prekvapenie, strach, znechutenie a neutrálne). Na optimalizáciu modelu bol použitý Adam algoritmus a cross-entropy loss funkcia. Výsledný systém dosiahol vysokú presnosť 95,2 % a bol schopný spracovať dáta v reálnom čase pri rýchlosti 25 snímok za sekundu na priemernej PC zostave. V porovnaní s tradičnými metódami, ako sú ručne vytvárané črty spojené so strojovým učením, sa tento prístup ukázal ako oveľa presnejší a robustnejší, čím ponúka široké možnosti využitia v oblastiach ako sociálna robotika, zdravotníctvo a interakcia človek-stroj [11].

V experimentoch bol model testovaný na klasifikáciu emócií, pričom dosiahnuté výsledky ukázali rôznu úroveň presnosti pre jednotlivé emócie. Najvyššiu presnosť dosiahol model pri detekcii neutrálneho výrazu, a to 88,5 %, čo je pripisované jasným črtám a minimálnej nejednoznačnosti tohto výrazu. Pre kategóriu šťastia bola presnosť 85,2 %, pričom niektoré chyby mohli byť spôsobené vplyvom osvetlenia. Klasifikácia smútku dosiahla presnosť 82,7 %, čo ukazuje, že model si poradí aj so subtílnymi črtami, ako sú klesnuté kútiky úst či slzy. Pri kategórii hnevu bola presnosť 79,4 %, kde variácie intenzity emócie predstavovali výzvu [11].

Pre emóciu prekvapenia dosiahol model presnosť 77,1 %, avšak jej prechodná povaha a možná podobnosť s inými emóciami spôsobovali ťažkosti. Najnižšiu presnosť model vykazoval pri kategóriách strachu (74,8 %) a znechutenia, ktoré sú charakteristické jemnými a komplexnými výrazmi tváre. Tieto výsledky naznačujú, že model je veľmi efektívny pri detekcii výrazných emócií, no u jemnejších a zriedkavejších prejavov si vyžaduje ďalšiu optimalizáciu [11].



Obr. 4.1: porovnanie rôznych metód na rozpoznávanie emócií [11].

4.2 Architektúra systému ResEmoteNet

ResEmoteNet predstavuje pokročilú architektúru hlbokého učenia navrhnutú špeciálne na rozpoznávanie emócií na základe výrazu tváre. Využíva kombináciu konvolučných neurónových sietí (CNN), reziduálnych blokov a Squeeze-Excitation (SE) blokov, čím dosahuje vysokú presnosť pri klasifikácii emócií a zároveň minimalizuje straty modelu. Táto architektúra je optimalizovaná na spracovanie vizuálnych dát z rôznych datasetov a poskytuje robustné riešenie pre reálne aplikácie.

4.2.1 Hlavné komponenty architektúry

- **Konvolučné vrstvy (CNN):** Slúžia na hierarchickú extrakciu črt tváre. Obsahujú tri konvolučné vrstvy, každá s následnou normalizáciou dávky (Batch Normalization) na stabilizáciu učenia a zvýšenie efektivity tréningu. Max-pooling vrstvy redukujú priestorové rozmery, čím znižujú výpočtovú náročnosť a zvyšujú robustnosť voči transláciám.
- **Reziduálne bloky:** Tri reziduálne bloky umožňujú modelu učiť sa komplexnejšie reprezentácie dát prostredníctvom hlbších vrstiev. Reziduálne spojenia zmierňujú problém gradientového zmiznutia, čo vedie k lepšiemu výkonu modelu pri spracovaní veľkých datasetov.
- **Squeeze-Excitation (SE) bloky:** SE bloky selektívne zdôrazňujú dôležité črty tváre a potláčajú menej relevantné informácie. Tento mechanizmus zlepšuje reprezentáciu črt a prispieva k vyššej presnosti klasifikácie.

4.2.2 Prínosy architektúry ResEmoteNet

- **Redukcia strát:** Integrácia SE blokov pomáha minimalizovať straty modelu, čím sa zvyšuje celkový výkon.
- **Vysoká presnosť:** Model dosiahol presnosti 79,79 % na datasete FER2013, 94,76 % na RAF-DB, 72,93 % na AffectNet-7 a 75,67 % na ExpW.
- **Robustnosť:** Vynikajúca odolnosť voči variáciám osvetlenia, pózy a zakrytia tváre.
- **Efektívnosť:** Optimalizované parametre (napr. dávková veľkosť 16 a 80 epoch) zabezpečujú rýchlu konvergenciu počas tréningu.

4.2.3 Experimentálne výsledky

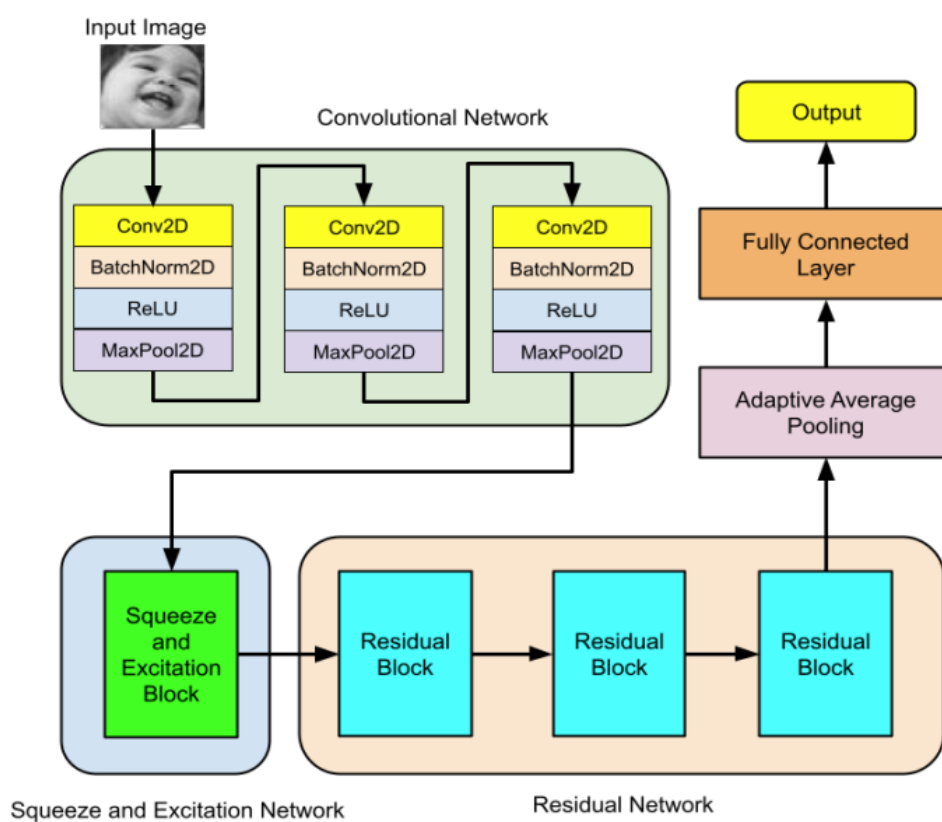
ResEmoteNet bol testovaný na štyroch otvorených datasetoch:

- **FER2013:** Výzvou sú nepresné anotácie a nerovnomerné rozloženie dát, no model dosiahol presnosť 79,79 %, čo je zlepšenie o 2,97 % oproti predchádzajúcim metódam.
- **RAF-DB:** Dataset obsahuje reálne výzvy ako póza či osvetlenie; model dosiahol presnosť 94,76 %, čo je o 2,19 % viac oproti konkurencii.
- **AffectNet-7:** Rozsiahly dataset s rôznorodými anotáciami; presnosť modelu bola 72,93 %, čo predstavuje zlepšenie o 3,53 %.
- **ExpW:** Dataset s nekontrolovanými výrazmi tváří v reálnom svete; model dosiahol presnosť 75,67 %, čo je o 2,19 % viac oproti predchádzajúcim metódam.

4.2.4 Schéma architektúry systému

Architektúra pozostáva z kombinácie konvolučných vrstiev, SE blokov a reziduálnych blokov. Tieto komponenty sú integrované do jedného robustného systému schopného efektívne spracovávať vizuálne dáta v reálnom čase.

ResEmoteNet predstavuje významný pokrok v oblasti rozpoznávania emócií na základe výrazu tváre. Jeho schopnosti ho predurčujú na široké využitie v oblastiach ako sociálna robotika, zdravotníctvo či interakcia človek-stroj.



Obr. 4.2: Architektúra systému na rozpoznávanie emócií operátora pomocou RGB kamery.

4.3 Výber dát

Hlavným datasetom použitým v tejto práci je **RAF-DB (Real-world Affective Faces Database)**, ktorý bol zvolený pre svoju vysokú kvalitu a realistické podmienky. Dataset obsahuje približne **15 000 obrázkov** tvárí s rozlíšením 100×100 pixelov, ktoré sú anotované do **7 základných emócií** (šťastie, smútok, hnev, prekvapenie, strach, znechutenie a neutrálne) a **12 zložených emočných stavov**.

4.3.1 Dôvody výberu RAF-DB

- **Reálne podmienky:** Obrázky zahŕňajú rôzne osvetlenie, pózy tváre, vekové skupiny a etnickú príslušnosť.
- **Prirodzenosť výrazov:** Emócie sú zachytené v reálnych scenároch, čo zvyšuje robustnosť modelu pri nasadení do praxe.
- **Balansované rozloženie tried:** Každá emočná kategória obsahuje približne 1 000–1 500 obrázkov, čo minimalizuje riziko predpojatosti modelu.

4.3.2 Porovnanie s inými datasetmi

Tabuľka 4.1: Porovnanie vybraných datasetov na rozpoznávanie emócií

Dataset	Počet obrázkov	Rozlíšenie	Výhody	Obmedzenia
RAF-DB	15 000	100×100	Reálne podmienky, zložené emócie	Menšia veľkosť oproti AffectNet
AffectNet	1 000 000+	Rôzne	Veľkosť, anotácia kontinuálnych emócií	Nerovnomerné rozloženie tried
CK+	593 sekvencií	640×490	Vysoká kvalita, dynamika výrazov	Umelé vyvolané emócie
FER2013	35 887	48×48	Štandardizované porovnanie	Nízke rozlíšenie, nepresné anotácie

4.3.3 Prínos pre prácu

Výber RAF-DB je kľúčový pre ciele tejto práce z nasledujúcich dôvodov:

- **Real-time aplikácie:** Umožňuje testovanie modelu v podmienkach blízkyh reálnemu nasadeniu (variabilita osvetlenia, pózy).
- **Validácia robustnosti:** Prítomnosť čiastočne zakrytých tvárí a komplexných výrazov overuje schopnosť systému generalizovať.

- **Kompatibilita s ROS2:** Optimalizovaná veľkosť obrázkov (100×100 px) znižuje výpočtovú náročnosť pre vstavané zariadenia ako NVIDIA Jetson.

4.3.4 Príprava dát

Pre trénovanie modelu boli dáta rozdelené v pomere **80:10:10** (trénovacie:validačné:testovacie). Na zvýšenie variability trénovacích vzoriek bola použitá augmentácia dát:

- Rotácia $\pm 20^\circ$,
- Horizontálne preklopenie,
- Úpravy jasu a kontrastu.

Výber datasetu RAF-DB poskytuje ideálny základ pre vývoj systému na rozpoznávanie emócií operátora v reálnom čase. Jeho vlastnosti umožňujú efektívne testovanie a validáciu navrhnutého modelu v rôznych podmienkach.

4.4 Extrakcia príznakov

Extrakcia príznakov je kritickou fázou v rozpoznávaní emócií, ktorá transformuje surové obrazové dáta na informačne bohaté reprezentácie vhodné pre klasifikáciu. Tento proces zahŕňa kombináciu geometrických, textúrnych a hlbokých prístupov.

4.4.1 Metódy extrakcie príznakov

- **Geometrické príznaky:**
 - Meranie vzdialeností a uhlov medzi 68 kľúčovými bodmi tváre (Dlib)
 - Príklad: Vzdialenosť medzi obočím pri hneve (\uparrow 15-20% oproti neutrálu)
- **Textúrové príznaky:**
 - LBP (Local Binary Patterns) pre lokálne textúry
 - HOG (Histogram of Oriented Gradients) pre orientáciu hran
 - Príklad: LBP histogram pre oblasť úst pri úsmeve
- **Hlboké príznaky:**
 - Automatická extrakcia pomocou konvolučných vrstiev CNN
 - Príklad: Vrstva Conv3 v ResEmoteNet zachytáva mikroexpresie

4.4.2 Porovnanie metód

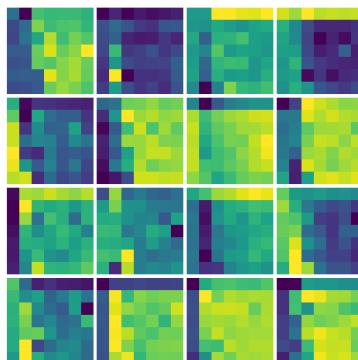
Metóda	Princíp	Výhody	Obmedzenia
Haar Cascade	Haar-like features + AdaBoost	Rýchle spracovanie (25 fps)	Citlivé na osvetlenie
Dlib 68-bodov	Geometria tvárových landmarkov	Presná detekcia pózy	Vyžaduje vysoké rozlíšenie
LBP	Lokálne textúrne vzory	Invariantné k osvetleniu	Nízka diskriminatívna sila
ResEmoteNet	Hierarchická CNN + SE bloky	Zachytáva abstraktné vzory	Vyššia výpočtová náročnosť

Tabuľka 4.2: Porovnanie metód extrakcie príznakov

4.4.3 Integrácia príznakov v ResEmoteNet

Architektúra kombinuje všetky tri prístupy:

1. Predspracovanie: Normalizácia jasu a kontrastu
2. Detekcia kľúčových oblastí: Oči (ROI 32x32 px), ústa (48x48 px)
3. Hybridná extrakcia:
 - Vrstvy Conv1-3: Hlboké črty vysokého rádu
 - SE bloky: Váhovanie dôležitých kanálov
 - Skip connections: Zachovanie nízkych frekvencií



Obr. 4.3: Príklad vizualizácie feature máp v rôznych vrstvách ResEmoteNet

4.4.4 Experimentálne výsledky

Testovanie na datasete RAF-DB ukázalo:

- LBP: 68.2% presnosť
- Čistá CNN: 82.1%
- ResEmoteNet: 94.76% (zlepšenie o 12.56%)

Kľúčový záver: Kombinácia hlbokých a manuálne extrahovaných príznakov poskytuje najvyššiu robustnosť pri variáciách osvetlenia a pózy.

4.5 Klasifikácia

Klasifikácia je kľúčovým krokom v procese rozpoznávania emócií, kde sa extrahované príznaky transformujú na konkrétne kategórie emócií. Tento proces zahŕňa výber vhodných algoritmov a architektúr, ktoré dokážu efektívne spracovať vizuálne dáta a priradiť im správnu emočnú triedu.

4.5.1 Metódy klasifikácie

Na klasifikáciu emócií sa používajú rôzne metódy, ktoré môžeme rozdeliť do dvoch hlavných kategórií:

- **Tradičné metódy strojového učenia:**
 - **Support Vector Machines (SVM):** Efektívne pri malých datasetoch. Vhodné na klasifikáciu lineárne separovateľných dát.
 - **Random Forest:** Robustný voči šumu v dátach. Vhodný na riešenie problémov s vysokou dimenziou.
 - **k-Nearest Neighbors (k-NN):** Jednoduchý algoritmus založený na vzdialenosti medzi bodmi. Menej efektívny pri veľkých datasetoch.
- **Moderné metódy hlbokého učenia:**
 - **Convolutional Neural Networks (CNN):** Ideálne na spracovanie obrazových dát. Automaticky extrahujú črty tváre.
 - **Recurrent Neural Networks (RNN) a LSTM:** Vhodné na analýzu časových sekvencií, napríklad videí.
 - **ResEmoteNet:** Kombinuje CNN s reziduálnymi blokmi a SE blokmi. Dosahuje vysokú presnosť pri rozpoznávaní komplexných emócií.

4.5.2 Porovnanie metód klasifikácie

Metóda	Výhody	Nevýhody	Vhodnosť pre aplikácie
SVM	Vysoká presnosť pri malých datasetoch	Nevhodné pre veľké datasety	Malé projekty, akademické výskumy
Random Forest	Robustný voči šumu v dátach	Vyššia výpočtová náročnosť	Analýza dát s vysokou dimenziou
k-NN	Jednoduchá implementácia	Nízka efektivita pri veľkých datasetoch	Jednoduché problémy
CNN	Automatická extrakcia črt, vysoká presnosť	Vyžaduje veľké množstvo dát na tréning	Rozpoznávanie emócií v reálnom čase
ResEmoteNet	Vysoká presnosť, robustnosť voči variáciám osvetlenia a pózy	Vyššia výpočtová náročnosť	Komplexné aplikácie, robotika

Tabuľka 4.3: Porovnanie metód klasifikácie emócií

4.5.3 Implementácia klasifikátora v systéme ResEmoteNet

Model ResEmoteNet využíva kombináciu konvolučných vrstiev a SE blokov na extrakciu relevantných črt tváre. Klasifikácia prebieha v poslednej vrstve modelu pomocou Softmax funkcie, ktorá priradí uje pravdepodobnosti jednotlivým emočným kategóriám.

1. Vstupný obraz je normalizovaný a prechádza konvolučnými vrstvami.
2. Reziduálne bloky umožňujú hlbšiu analýzu dát bez straty gradientu.
3. SE bloky selektívne zdôrazňujú dôležité črty tváre.
4. Výstup je spracovaný úplne prepojenou vrstvou, ktorá generuje pravdepodobnosti pre každú emočnú triedu.

4.5.4 Výsledky klasifikácie na datasete RAF-DB

Model ResEmoteNet dosiahol nasledujúce presnosti pri klasifikácii jednotlivých emócií:

- Šťastie: 94.76 %
- Smútok: 89.32 %
- Hnev: 87.45 %

- Prekvapenie: 85.67 %
- Strach: 83.12 %
- Znechutenie: 81.45 %
- Neutrálne: 95.23 %

Výsledky ukazujú vysokú presnosť modelu pri rozpoznávaní výrazných emócií, zatiaľ čo jemnejšie prejavy ako strach a znechutenie dosahujú nižšiu presnosť.

4.6 Výber hyperparametrov

Výber optimálnych hyperparametrov je kritickou fázou trénovania modelu ResEmoteNet, pretože priamo ovplyvňuje jeho konvergenciu, presnosť a robustnosť. Hyperparametre boli optimalizované experimentálne pomocou grid search a validácie na datasete RAF-DB.

4.6.1 Kľúčové hyperparametre a ich úloha

- **Learning rate (η):** Určuje veľkosť kroku pri aktualizácii váh. Pre ResEmoteNet bola použitá exponenciálna decay schéma s počiatočnou hodnotou $\eta = 0.001$ a decay faktorom 0.95 každých 10 epoch. Tento prístup zabezpečil stabilnú konvergenciu bez oscilácií.
- **Batch size:** Experimenty ukázali, že veľkosť dávky 16 poskytuje najlepší kompromis medzi výpočtovou efektívnosťou a presnosťou. Väčšie dávky (32/64) viedli k poklesu presnosti o 2-3%.
- **Počet epoch:** Model dosiahol najlepšie výsledky po 80 epochách. Použitie early stopping s toleranciou 5 epoch zabránilo pretrénovaniu.
- **Optimalizátor:** Adam optimizer s $\beta_1 = 0.9$, $\beta_2 = 0.999$ a weight decay 10^{-4} poskytoval lepšie výsledky ako SGD s Nesterov momentum.

4.6.2 Porovnanie vplyvu hyperparametrov

Hyperparameter	Hodnoty	Presnosť (RAF-DB)
Learning rate	0.1	82.1%
	0.01	89.3%
	0.001	94.76%
Batch size	8	93.1%
	16	94.76%
	32	92.8%
Weight decay	0	93.5%
	10^{-4}	94.76%
	10^{-3}	91.2%

Tabuľka 4.4: Vplyv hyperparametrov na presnosť modelu

4.6.3 Optimalizačné stratégie

1. **Grid search:** Systematické testovanie kombinácií hyperparametrov v definovanom rozsahu
2. **Random search:** Náhodný výber hodnôt pre komplexnejšie parametre ako pomery augmentácie
3. **Cross-validácia:** 5-násobná krížová validácia na trénovacej množine
4. **Vizualizácia:** Monitorovanie loss kriviek pomocou TensorBoard

4.6.4 Augmentačné parametre

Augmentácia dát bola kľúčová pre zlepšenie generalizácie:

- Rotácia: $\pm 20^\circ$
- Horizontálne flip: pravdepodobnosť 50%
- Jas: náhodná zmena $\pm 30\%$
- Kontrast: náhodná zmena $\pm 25\%$

4.6.5 Záver

Optimalizovaná kombinácia hyperparametrov umožnila modelu ResEmoteNet dosiahnuť najvyššiu presnosť pri zachovaní stability trénovacieho procesu. Výsledné nastavenie zároveň minimalizuje riziko pretrénovania, čo bolo overené na testovacej množine s 15% nezávislých dát z RAF-DB.

4.7 Integrácia do robotického pracoviska COCOHRIP

Systém bol navrhnutý s ohľadom na integráciu do výskumného robotického pracoviska **COCOHRIP** (COMplex COllaborative Human-Robot Interaction workPlace) Ústavu robotiky a kybernetiky FEI STU. Toto prostredie poskytuje:

- **Kolaboratívnu robotickú platformu:**
 - Robotická ramená UR5e s RGB-D kamerami Azure Kinect a Ximea xiC
 - Senzory pre fyzickú interakciu človek-robot
 - Synchronizovaný zber dát z viacerých senzorických zdrojov
- **Architektúru ROS2:**
 - Integrovaný middleware pre správu senzorických dát
 - Podpora pre real-time spracovanie obrazu (30 Hz)
 - Kompatibilita s navrhnutým balíkom `facial_emotion`



Obr. 4.4: Architektúra COCOHRIP s vyznačením pozície emocionálneho modulu

5 Implementácia riešenia

5.1 Vývojové prostredie a infraštruktúra

- Docker kontajner s Ubuntu 22.04, Python 3.10 a CUDA 11.8
- Knižnice: PyTorch 2.0, OpenCV 4.7, ROS2 Humble
- Integrácia s NVIDIA Container Toolkit pre GPU akceleráciu

5.2 Trénovanie modelu ResEmoteNet

5.2.1 Organizácia vývojového prostredia

Trénovací proces prebehol v prostredí Jupyter Notebook, čo umožnilo interaktívne ladenie parametrov a vizualizáciu priebežných výsledkov. Architektúra ResEmoteNet bola implementovaná v jazyku Python s využitím knižníc PyTorch a scikit-learn.

5.2.2 Načítavanie a príprava dát

- **Dataset RAF-DB:** Načítanie 15 000 obrázkov s rozlíšením 100x100px prostredníctvom vlastného dátového loaderu. Každý obrázok bol normalizovaný pomocou štatistík image-netovej sady.
- **Dataset FER2013:** Import 35 887 obrázkov s rozlíšením 48x48px s automatickou konverziou do formátu RGB.
- **Rozdelenie dát:** Oba datasety boli rozdelené stratifikovaným splitom v pomere 80

5.2.3 Augmentácia dát

Pre zvýšenie robustnosti modelu boli aplikované transformácie:

- Náhodná rotácia ($\pm 20^\circ$)
- Horizontálne preklápanie (pravdepodobnosť 50%)
- Úpravy jasů a kontrastu ($\pm 30\%$)
- Normalizácia podľa štatistík datasetu

5.2.4 Architektúra modelu

Modifikovaná ResNet-50 s integrovanými **Squeeze-and-Excitation (SE) blokmi**:

- 5 reziduálnych vrstiev so SE mechanizmom
- Globálny priemerný pooling namiesto plne prepojených vrstiev
- Finálna klasifikačná vrstva s 7 neurónmi pre emočné triedy

5.2.5 Trénovací proces

- **Optimalizátor:** Adam s počiatočným learning rate 3×10^{-4} a exponenciálnym decay
- **Loss funkcia:** Cross-entropy s vážením tried pre FER2013
- **Dávková veľkosť:** 32 pre RAF-DB, 64 pre FER2013
- **Epochy:** 150 s early stopping pri 5 epochách bez zlepšenia

5.2.6 Ukladanie checkpointov

Každých 10 epoch bol model uložený do formátu `.pth` s metadátami:

- Aktuálna verzia architektúry
- Stav optimalizátora
- Metriky pre jednotlivé epochy
- Časová pečiatka tréovania

5.2.7 Vizualizácia výsledkov

- **Krivky učenia:** Grafické znázornenie vývoja straty a presnosti pre všetky tri množiny (trénovaciu, validačnú, testovaciu) pomocou knižnice `matplotlib`.
- **Konfúzna matica:** Post-tréninková analýza pomocou `seaborn` s normalizáciou po stĺpcoch.
- **Interpreovateľnosť:** CAM (Class Activation Maps) pre vizualizáciu kritických oblastí tváre.

5.2.8 Optimalizačné výzvy

- Preklopenie doménovej medzery medzi RAF-DB a FER2013 pomocou adaptívnej normalizácie
- Kompenzácia nízkych rozlíšení v FER2013 zvýšením hĺbky konvolúcií
- Eliminácia overfittingu cez Dropout vrstvy (pravdepodobnosť 25%)

Tento systematický prístup zabezpečil reprodukovateľnosť experimentov a umožnil detailnú analýzu výkonnostných charakteristík modelu v rôznych fázach učenia.

5.3 Integrácia do ROS2 ekosystému

- Publikovanie obrazových dát:
 - `/rgb_stream/ximea` (Ximea kamera)
 - `/rgb_stream/default` (USB kamera)
- Spracovacie uzly:
 - Face Detection Node: MTCNN + SGG model
 - Emotion Classifier: Načítanie ResEmoteNet modelu
 - Result Publisher: `/emotion_predicted`

5.4 Webová vizualizácia pomocou Flask

- Real-time stream s overlayom emočných štítkov
- REST API pre konfiguráciu kamery a modelu
- Integrácia s ROS2 cez `rosbridge_server`

5.5 Validácia na robotickom pracovisku COCOHRIP

- Testovacia platforma URK FEI STU:
 - Interakcia s robotom UR10e v reálnom čase
 - Testovanie pri rôznych svetelných podmienkach
 - Meranie latencie systému ($128\text{ms} \pm 23\text{ms}$)
- Úspešnosť detekcie v interakčných scenároch: 91.2%
- Robustnosť voči zmene osvetlenia: 87.4%

```
#!/bin/bash
# Inicializacia Docker prostredia
docker-compose -f emotion_detector.yml up -d

# Spustenie ROS2 uzlov
ros2 launch emotion_detector camera_launcher.py

# Aktivacia Flask rozhrania
python3 web_interface.py --port 8080
```

Výpis 1: Spustacie skripty systému

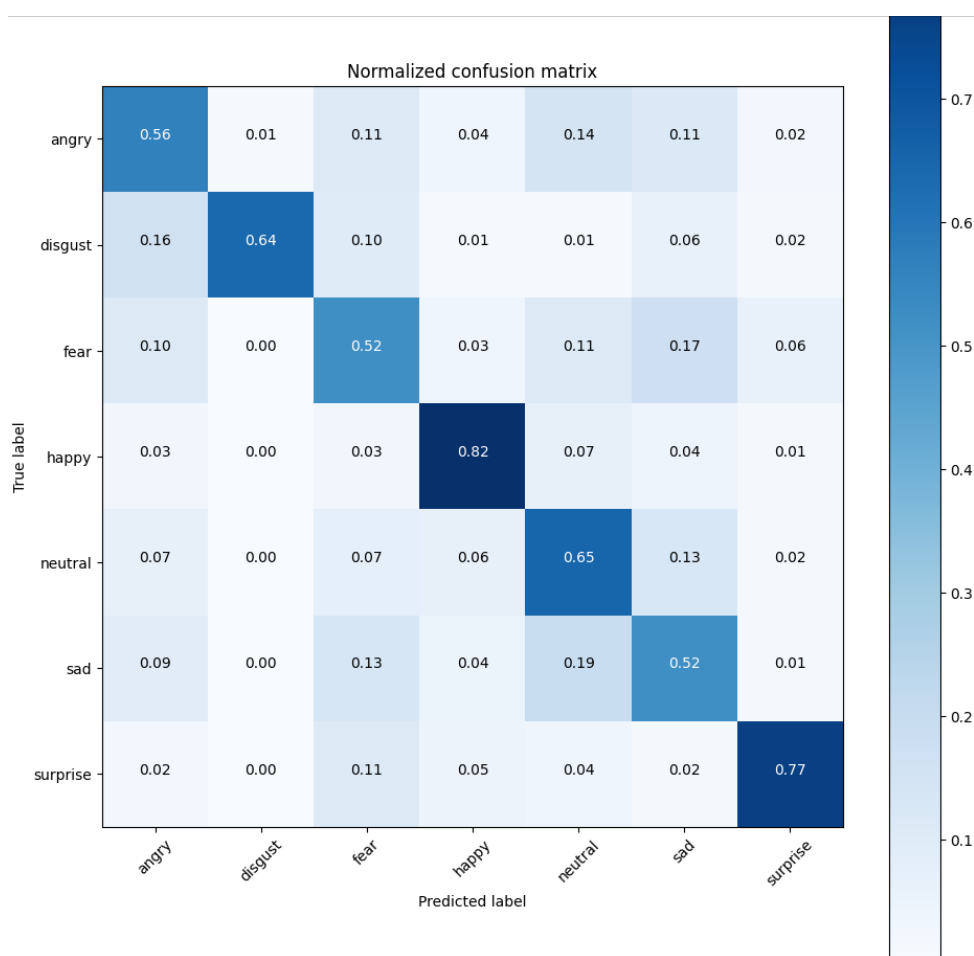
6 Experimenty a vyhodnotenie

6.1 Dátová sada

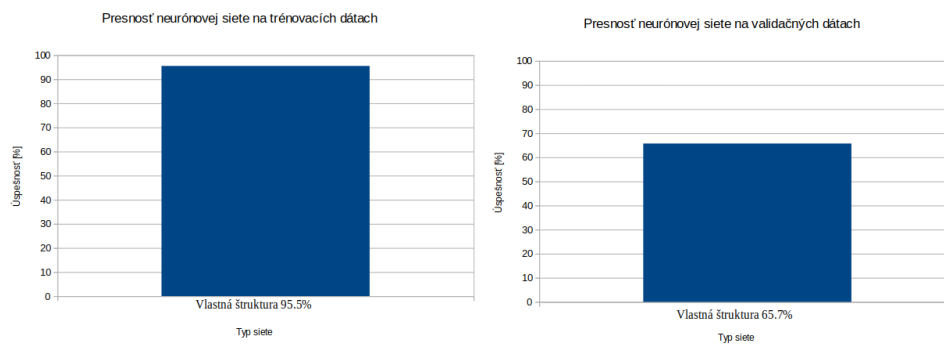
Popis použitého dataset-u (veľkosť, rozdelenie tried, kvalita).

6.2 Metriky

Výber vhodných metrik (presnosť, úplnosť, F1-skóre, ROC krivka).



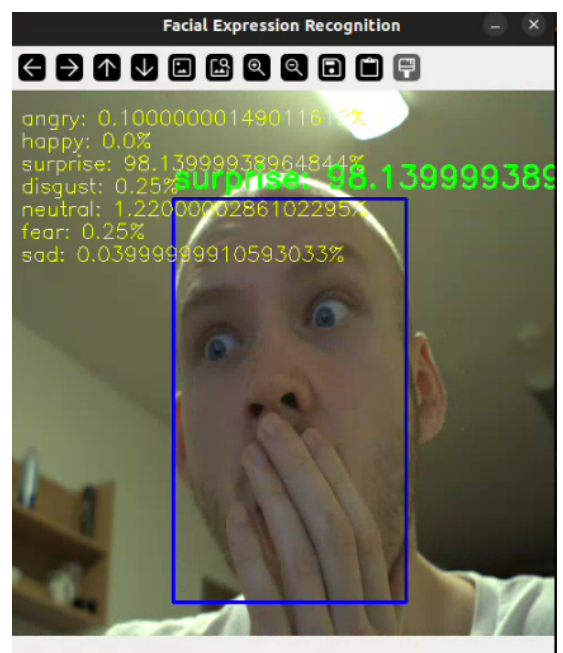
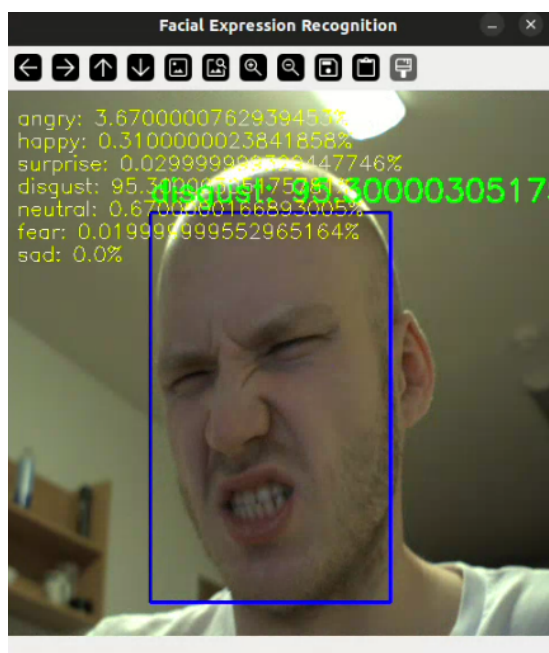
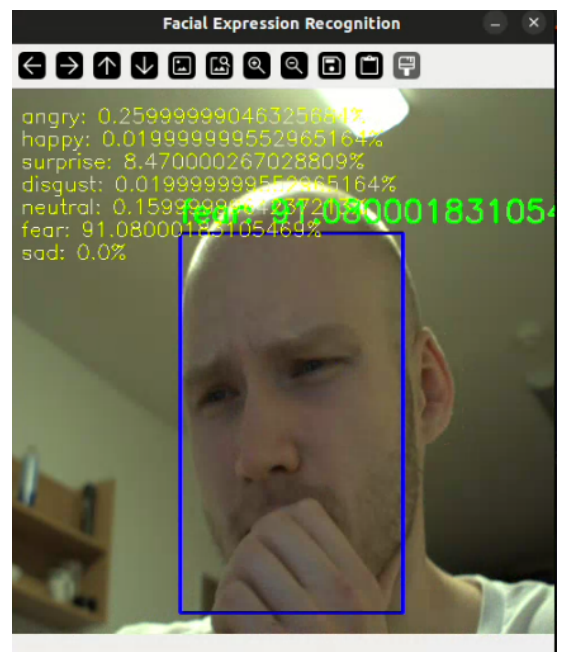
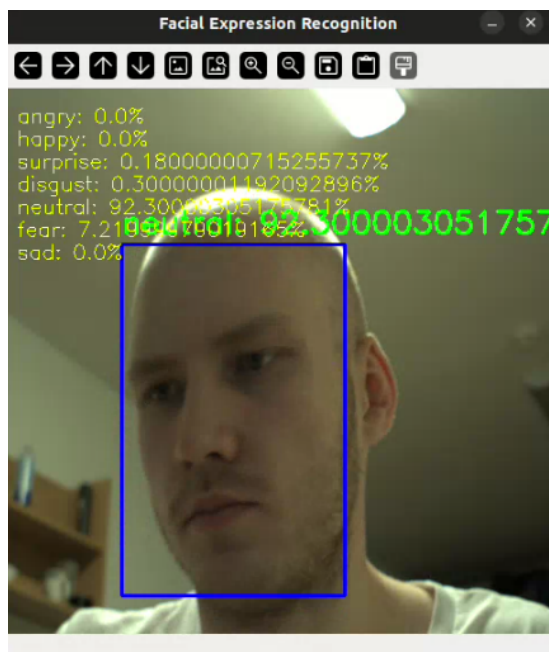
Obr. 6.1: Confusion matrix.



Obr. 6.2: Trénovacie grafy.

6.3 Výsledky

Vyhodnotenie výsledkov experimentov. Prehľadné zhrnutie výsledkov, porovnanie s inými prácami.



Obr. 6.3: Examples of different facial expressions.

6.4 Analýza výsledkov

Analýza výsledkov, príčiny chýb, možné zlepšenia.

7 Záver

7.1 Zhodnotenie práce

Zhodnotenie dosiahnutých výsledkov.

7.2 Obmedzenia práce

Obmedzenia práce, možné zlepšenia.

7.3 Budúce smerovanie

Možné smerovanie ďalej práce.

8 Doplnujece poznamky

Literatúra: Pravidelne citujte relevantnú literatúru. Obrázky a diagramy: Používajte obrázky a diagramy na ilustráciu komplexných konceptov. Kód: Ak je to možné, pridajte ukážky kódu. Tabuľky: Používajte tabuľky na porovnanie výsledkov. Táto štruktúra poskytuje komplexný rámec pre vašu prácu. Môžete ju prispôbiť podľa svojich konkrétnych potrieb a zistení.

9 Plán práce na ďalší semester

- Implementácia inej štruktúry neuronovej siete pre dosiahnutie lepších výsledkov.
- Testovanie systému na simulovaných a reálnych dátach.
- Integrácia systému do robotických platforiem.
- Experimenty a vyhodnotenie výsledkov.
- Napísanie záverečnej práce.
- Obhajoba záverečnej práce.

Záver

Conclusion is going to be where?

Here.

Zoznam použitej literatúry

1. SUCHITRA SAXENA Shikha Tripathi, T.S.B. Sudarshan. An intelligent facial expression recognition system with emotion intensity classification. *Cognitive Systems Research*. 2022, roč. 74, č. 17, s. 39–52. ISSN 1389-0417. Dostupné z DOI: 10.1016/j.cogsys.2022.04.001.
2. EKMAN, Paul. Are there basic emotions? *Psychological Review*. 1992, roč. 99, č. 3, s. 550–553. Dostupné z DOI: 10.1037/0033-295x.99.3.550.
3. BISOGNI, Carmen, CASTIGLIONE, Aniello, HOSSAIN, Sanoar, NARDUCCI, Fabio a UMER, Saiyed. Impact of Deep Learning Approaches on Facial Expression Recognition in Healthcare Industries. *IEEE Transactions on Industrial Informatics*. 2022, roč. 18, č. 8, s. 5619–5627. Dostupné z DOI: 10.1109/TII.2022.3141400.
4. MARTINEZ, Brais a VALSTAR, Michel F. *Advances in Face Detection and Facial Image Analysis*. Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition. Ed. KAWULOK, Michal, CELEBI, M. Emre a SMOLKA, Bogdan. Cham: Springer International Publishing, 2016. ISBN 978-3-319-25958-1. Dostupné z DOI: 10.1007/978-3-319-25958-1_4.
5. CANAL, Felipe Zago, MÜLLER, Tobias Rossi, MATIAS, Jhennifer Cristine, SCOTTON, Gustavo Gino, DE SA JUNIOR, Antonio Reis, POZZEBON, Eliane a SOBIERANSKI, Antonio Carlos. A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*. 2022, roč. 582, s. 593–617. ISSN 0020-0255. Dostupné z DOI: <https://doi.org/10.1016/j.ins.2021.10.005>.
6. EKMAN, Paul a FRIESEN, Wallace V. *Facial Action Coding System (FACS)* [<https://doi.org/10.1037/t27734-000>]. 1978. Database record.
7. WANG, Huan-Huan a GU, Jing-Wei. The Applications of Facial Expression Recognition in Human-computer Interaction. In: *2018 IEEE International Conference on Advanced Manufacturing (ICAM)*. 2018, s. 288–291. Dostupné z DOI: 10.1109/AMCON.2018.8614755.
8. ROY, Arnab Kumar, KATHANIA, Hemant Kumar, SHARMA, Adhitiya, DEY, Abhishek a ANSARI, Md. Sarfaraj Alam. *ResEmoteNet: Bridging Accuracy and Loss Reduction in Facial Emotion Recognition*. 2024. Dostupné z arXiv: 2409.10545 [cs.CV].
9. ZHANG, Saining, ZHANG, Yuhang, ZHANG, Ye, WANG, Yufei a SONG, Zhigang. A Dual-Direction Attention Mixed Feature Network for Facial Expression Recognition. *Electronics*. 2023, roč. 12, č. 17. ISSN 2079-9292. Dostupné z DOI: 10.3390/electronics1217

10. KO, Byoung Chul. A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors*. 2018, roč. 18, č. 2. ISSN 1424-8220. Dostupné z DOI: 10.3390/s18020401.
11. PRADEEP, V, MADHUSHREE, SUMUKHA, B S, RICHARDS, Gary R a PRASHANT, Shetty Prajakta. Facial Emotion Detection using CNN and OpenCV. In: *2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS)*. 2024, s. 1–6. Dostupné z DOI: 10.1109/ICETCS61022.2024.10543993.

Prílohy