Prepoznavanje emocija na ljudskom licu

Radić, Igor

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA

I INFORMACIJSKIH TEHNOLOGIJA

Sveučilišni studij

PREPOZNAVANJE EMOCIJA NA LJUDSKOM LICU

Diplomski rad

Igor Radić

Osijek, 2023.

SADRŽAJ

1. UVOD

Način komunikacije između ljudi u svakodnevnom životu ovisi, između ostalog, i o emocijama pod čijim utjecajem su sudionici razgovora. Ton i ugođaj izgovorenih ili napisanih riječi kao i odabir riječi uvelike ovisi o emociji čovjeka kojem se želi uputiti neka poruka ili informacija. Shodno tome radi li se o interakciji računala i čovjeka, dobro je da računalo ima neki način zapažanja emocije korisnika koji ga trenutno koristi kako bi mogao prilagoditi sadržaj koji pruža korisniku, prilagoditi izgled sučelja ili kako bi mogao djelovati na neki drugi način na korisnika ukoliko se radi o nekom ugradbenom računalnom sustavu kojemu je bitno emocionalno stanje korisnika. Emocije se kod ljudi učinkovito mogu prepoznati promatrajući izraz lica, stoga računalni sustavi mogu prepoznavati emocionalno stanje korisnika putem kamere koja snima korisnika. Ovakvo zapažanje emocija korisnika široko je primjenjivo jer današnji pametni telefoni, tableti i prijenosna računala imaju kameru. Neki od problema kod ovog pristupa mogu biti osvjetljenje, različite udaljenosti lica od kamere, različite pozicije lica na slici ali i strah korisnika od zloupotrebljavanja prikupljenih slika zbog kojeg bi korisnik mogao onemogućiti uzorkovanje slika potrebnih za prepoznavanje emocija korisnika. No unatoč navedenim problemima, računalna rješenja za prepoznavanje emocija sa ljudskog lica vrlo su zanimljiva jer bi ona mogla poboljšati interakciju čovjeka i računala. Rješenja koja se bave ovim problemom najčešće se baziraju na strojnom učenju ili nekoj od metoda računalnog vida.

U sklopu ovoga rada potrebno je istražiti područje prepoznavanja emocija na temelju slike ljudskog lica te pružiti uvid u glavne značajke ovog problema. Nakon toga potrebno je usporediti neka od najnovijih dostupnih (eng. *state-of-the-art*) rješenja i prikazati rezultate uspoređivanja i evaluacije. Na kraju rada, cilj je izraditi jednostavnu aplikaciju za prepoznavanje emocija na temelju slike ljudskog lica pomoću dostupnog rješenja koje se pokazalo kao najbolje. Referentni skup podataka na temelju kojega će biti vrednovana pojedina rješenja je FER-2013 [1]. Emocije koje će se prepoznavati u ovom radu su ljutnja, gađenje, strah, radost, tuga, iznenađenost te neutralnost.

Rad se sastoji od uvoda, pregleda područja teme gdje će biti opisan problem prepoznavanja emocije na temelju slike ljudskog lica kao i opis nekih od najnovije dostupnih skupova podataka i rješenja za ovaj problem. Potom slijedi opis praktičnog dijela rada u kojem se nalazi opis primjene nekih od dostupnih rješenja te pregled konačne aplikacije za prepoznavanje emocija na temelju slike ljudskog lica. Na samom kraju nalazi se zaključak koji sadrži glavna zapažanja i zaključke ovoga rada.

2. PREGLED PODRUČJA TEME

2.1. Opis problema

Ljudima je informacija o emocionalnom stanju čovjeka sa kojim komuniciraju vrlo bitna, ta informacija može se saznati putem verbalne i neverbalne komunikacije. Istraživanja su još šezdesetih godina prošloga stoljeća pokazala kako je 7% komunikacije verbalno, 38% vokalno te 55% vizualno [2]. Ova činjenica objašnjava želju za prepoznavanjem emocionalnog stanja čovjeka baš na temelju izraza lica. Ljudi doživljavaju velik broj emocija no sve se one mogu svrstati u neku od šest glavnih emocija, a to su sreća, tuga, iznenađenost, ljutnja, gađenje i strah [3]. U tablici 2.1. dan je opis izgleda lica za svaku od glavnih emocija prema [3].

Tab. 2.1. Opis izgleda lica za pojedinu emociju

Emocija	**Izgled lica**
Iznenađenost	Podignute i zakrivljene obrve, duge horizontalne bore na čelu, široko otvorene oči, otvorena usta, spuštena brada
Strah	Ravno podignute obrve, kratke bore na čelu, otvorene oči, napetost u donjim kapcima, usta mogu biti otvorena ili zatvorena
Sreća	Nema značajnih očitovanja na čelu i obrvama, donji kapci mogu biti podignuti što uzrokuje sužen izgled očiju, bore oko očiju, obrazi i rubovi usana podignuti prema gore, usta mogu biti zatvorena ili otvorena te se mogu vidjeti zubi
Tuga	Blage bore na sredini čela, unutarnji rubovi obrva podignuti, vanjski rubovi obrva spušteni, blago zatvorene oči, rubovi usana spušteni prema dolje
Ljutnja	Izraženije vertikalne bore na sredini čela, unutarnji vrhovi obrva spušteni prema dolje, vanjski rubovi obrva usmjereni prema gore, gornji kapci spušteni, donji kapci ponekad podignuti, usne snažno pritisnute ili otvorene te četvrtastog oblika, zubi se mogu vidjeti
Gađenje	Obrve spuštene, bore na spoju nosa i čela, donji kapci podignuti, podignuti obrazi, gornja usna podignuta, donja usna usmjerena prema naprijed i/ili prema van, može se vidjeti jezik blizu usana, u slučaju da su usta zatvorena rubovi usana su blago usmjereni prema dolje

Hipoteza dostupnih skupova podataka jest da ukoliko u određenom trenutku na licu promatrane osobe nisu prisutni izrazi lica opisani u tablici 2.1, tada kod te osobe nisu niti prisutne te emocije pa se može smatrati da je osoba trenutno neutralna po pitanju emocije. Više o ovome će biti u sljedećem potpoglavlju.

Prepoznavanje emocija na temelju izraza ljudskog lica problem je u kojem je potrebno odrediti emociju prisutnu kod čovjeka, čije lice se promatra, na temelje karakteristika opisanih u tablici 2.1. Iako je čovjeku taj problem relativno lak, računalo mora učiniti nekoliko koraka kako bi riješilo ovaj problem. Prvi korak je dohvaćanje slike čovjeka, nakon toga potrebno je na slici pronaći područje od interesa (eng. *region of interest*) što je u ovom slučaju lice, iz tog područja od interesa se izvlače odnosno određuju potrebne značajke te se pomoću nekog algoritma ili metode dolazi do rješenja tj. emocije koja je očitovana na licu osobe koja se promatra. U ovome radu obrađivat će se samo metode za prepoznavanje emocije iz slike na kojoj se nalazi samo lice, odnosno problem detekcije lica neće se obrađivati.

Jedno manje istraživanje koje je provedeno 2013. godine tvrdi da je ljudska točnost prepoznavanja emocija na temelju ljudskog lica 65±5 % [4]. Iz navedenog se može zaključiti kako na prepoznavanje emocija na temelju ljudskog lica utječe i subjektivan dojam osobe što govori o težini problema kojim se bavi ovaj rad budući da računala nemaju subjektivan dojam.

Čovjek može potisnuti ili glumiti neku od emocija no to nije uzeto u obzir prilikom proučavanja ovoga problema jer bi dodatno zakompliciralo već dovoljno kompleksan problem.

Iako je problem kompleksan, ova problematika je zanimljiva jer ima dosta primjene u područjima kao što su sigurnosni sustavi u vožnji, edukacija, medicina, zabava, interakcija čovjeka sa računalom (eng. *HCI – Human Computer Interaction*), socijalni roboti, virtualna asistencija, oglašavanje, zadovoljstvo korisnika, zadovoljstvo radnika, zabava i poboljšanje kvalitete života [5-7].

2.2. Skupovi podataka

U ovome potpoglavlju se nalazi kratak pregled nekih od najčešće korištenih skupova podataka za prepoznavanje emocija na temelju slike ljudskog lica. Dijagramom na slici 2.1. prikazani su odnosi brojeva znanstvenih radova, u kojima su citirani skupovi podataka koji su opisani u ovome poglavlju, kako bi se stekao dojam o njihovom korištenju u posljednjih pet godina.

Slika 2.1. Citiranost skupova podataka u posljednjih pet godina [8]

2.2.1. AffectNet

AffectNet je trenutno najveća baza podataka izraza lica, valencije i uzbuđenja koja omogućuje istraživanje automatiziranog prepoznavanja izraza lica. Sadrži više od 1 000 000 slika od kojih su ~ 440 000 slika ručno označene i podijeljene u klase od strane stručnih ljudi. Slike su preuzete sa interneta unošenjem 1 250 riječi (na 6 različitih jezika), koje se odnose na neku od emocija, u tri vodeće tražilice [9]. Prednost ovog skupa podataka je i ta što slike nisu nastale u kontroliranim uvjetima što ga čini vrlo upotrebljivim u stvarnom svijetu. Slika 2.2. daje uvid u raznovrsnost slika u ovome skupu podataka. AffectNet nije javno dostupan, moguće ga je dobiti isključivo u svrhe istraživanja uz vrlo stroge uvijete korištenja. Iako je ovo vjerojatno najbolji skup podataka za primjenu u području ovoga rada, kako nije javno dostupan nije upotrijebljen kao referentni skup u ovome radu.

Slika 2.2. Primjeri slika iz AffectNet skupa podataka [9]

2.2.2. CK+

CK+ popularan je i skraćeni naziv koji predstavlja *Extended Cohn-Kanade* skup podatak. Radi se o još jednom često korištenom skupu podataka u području prepoznavanja emocija na temelju slike ljudskog lica. Skup se sastoji od 593 slike 123 različita čovjeka. Subjekti koji su na slikama su u rasponu od 18 do 50 godina, 69 % subjekata su žene a 31 % muškarci. Slike su dimenzija 640x490 i 640x480 piksela, većina slika je crno-bijela no neke su i u boji. CK+ je skup podataka koji je nastao u kontroliranim uvjetima, to jest pozadina, osvjetljenje i poza ljudi koji su na slikama je identična [10]. Slike su raspoređene u 8 klasa, a to su ljutnja, gađenje, strah, sreća, tuga, iznenađenost, prezir i neutralnost. Slike nisu ujednačene što znači da broj slika u svakoj od klasa nije jednak. CK+ je javno dostupan skup podataka. Na slici 2.3. se mogu vidjeti neke od slika iz CK+ skupa podataka. Odmah na prvi pogled vidljivo je da na slikama ne dominiraju samo lica već je prisutan i značajan dio pozadine.

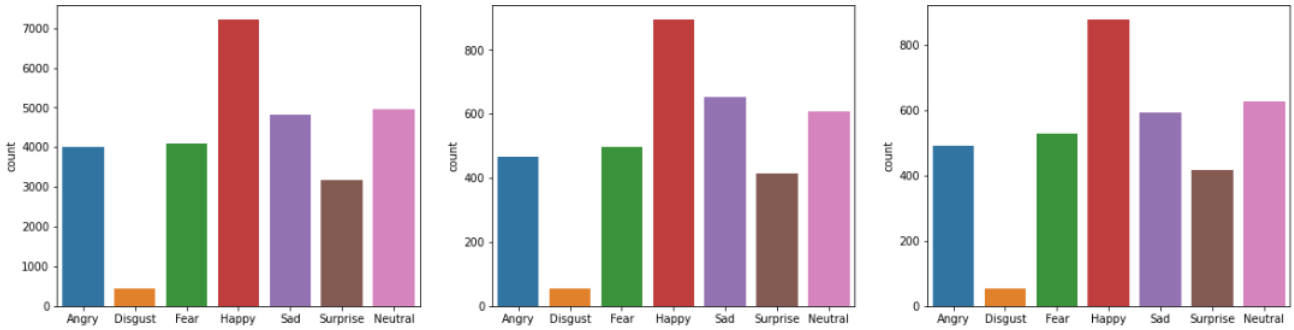
Slika 2.3. Primjeri slika iz CK+ skupa podataka [10]

2.2.3. FER-2013

Ovaj je skup podataka nastao 2013. godine za potrebe natjecanja u području prepoznavanja emocija na temelju slike ljudskog lica održanom na *International Conference on Machine Learning* iste godine. Slike su prikupljene pomoću Google-ovog API-ja za pretraživanje slika unoseći skup od približno 600 nizova riječi koji su nastale kombinacijom 184 riječi povezane sa nekom od emocija i riječi koje su povezane sa spolom, dobi i etičkom pripadnošću. Prvih 1000 slika, koje je pretraživanje dalo kao rezultat za pojedini niz riječi, pohranjeno je te je na njima vršena detekcija lica kao i izdvajanje područja lica sa slike. Tako izdvojene slike lica su potom pregledane od strane ljudi koji su neupotrebljive slike uklanjali a upotrebljive obrađivali te pohranjivali. Konačno, svim slikama je

promijenjena veličina na dimenziju 48x48 piksela. FER-2013 skup podataka sadrži 35 887 crno-bijelih slika koje su podijeljene u 7 klasa: sreća, tuga, ljutnja, gađenje, neutralnost, iznenađenost i strah. Za treniranje je predviđeno 28 709 slika, za validaciju 3 589 slika i za testiranje također 3 589 slika [4].

S obzirom na način nastajanja ovog skupa podataka, slike su doista raznovrsne što ovaj skup podataka čini bližim uvjetima u stvarnom svijetu. No ovakav način prikupljanja podataka ima i mana, jedna od kojih je ta što je broj slika u nekim klasama drastično veći ili manji u odnosu na broj slika u ostalim klasama. Dijagrami na slici 2.4. prikazuju odnos broja slika po klasama. Na dijagramu je vidljivo da u sva tri skupa prevladava broj slika koje spadaju u klasu sreća, dok je slika koje spadaju u klasu gađenje izrazito manje u usporedbi sa brojem slika u ostalim klasama.



a) b) c)

Slika 2.4. Odnos broja slika po klasama: a) Skup za treniranje b) skup za testiranje c) skup za validaciju [11]

Još jedna mana automatskog prikupljanja slika jesu neispravno klasificirane slike. Slika 2.5 prikazuje neke od neispravno klasificiranih slika.

~~sad~~ neutral | ~~angry~~ sad | ~~fear~~ angry | ~~neutral~~ sad | ~~surprise~~ fear

Slika 2.5. Prikaz neispravno klasificiranih slika u FER-2013 skupu podataka [1]

Ove mane ispravljene su u FER+ [12] skupu podataka no FER-2013 je, unatoč ovim manama, i dalje popularniji i korišteniji skup podataka od FER+ skupa podataka. Bitno je još spomenuti da je ovaj skup podataka javno dostupan.

S obzirom na javnu dostupnost ovog skupa podataka, njegovu popularnost, veliki broj slika koje sadrži te činjenicu da su slike vrlo raznovrsne, ovaj skup podataka koristi se kao referentni skup podataka u ovome radu te se na temelju njega uspoređuju rezultati pojedinih dostupnih rješenja.

2.3. Pregled postojećih rješenja

Rješenja za problem prepoznavanja emocija na temelju slike ljudskog lica mogu se svrstati u dvije kategorije, ona rješenja koja značajke iz slika izvlače pomoću nekog od algoritama strojnog učenja i ona rješenja koja značajke izvlače „ručno“, npr. pomoću neke od metoda računalnog vida kao što su detekcija rubova, detekcija bridova i slično. Znajući to, postavlja se pitanje koja kategorija rješenja daje bolje rezultate. Odgovor na to pitanje zanimao je i organizatore *International Conference on Machine Learning* 2013. godine kada su organizirali natjecanje u prepoznavanju emocija na temelju slike ljudskog lica te kreirali FER-2013 skup podataka za potrebe natjecanja [4]. Najbolje rješenje na natjecanju ostvarilo je točnost od 71.16 %. Navedeno rješenje koristi konvolucijsku neuronsku mrežu ali u zadnjem sloju ne sadrži *softmax* sloj već linearni stroj s potpornim vektorima (eng. *Linear SVM - Linear Support Vector Machine*) [13]. Prve tri najbolje metode na spomenutom natjecanju koriste konvolucijske mreže odnosno rješenja koja koriste automatski dobivene značajke dok četvrto najbolje rješenje koristi „ručno“ kreirane značajke iz čega se može zaključiti da rješenja koja za izvlačenje značajki iz slika koriste strojno učenje daju bolje rezultate od onih rješenja koja značajke izvlače „ručno“ [10].

Trenutno, također, veći broj rješenja za problem prepoznavanja emocije na temelju slike ljudskog lica primjenjuju konvolucijske neuronske mreže [14 - 19]. Postoje i rješenja koja koriste oba pristupa određivanja značajki. Jedno takvo rješenje opisano je u [20] a koristi se kombinacijom ručno dobivenih značajki pomoću SIFT detektora te automatski dobivenih značajki iz konvolucijskih mreža dok za klasifikaciju koristi stroj s potpornim vektorima.

Točnosti suvremenih rješenja na FER-2013 skupu podataka nešto su bolje od onih na natjecanju u sklopu *International Conference on Machine Learning* 2013. godine kada je spomenuti skup podataka i nastao, no ne znatno. Jedna od metoda kojom se nastoji poboljšati rezultat je kombinacija više

modela, na taj način su [14 - 16] uspjeli ostvariti točnosti veće od 75 %. Kod ovakvog pristupa za isti problem trenira se više modela te se konačna predikcija klase vrši uzimajući u obzir predikcije svih modela u kombinaciji. Mana ovakvih rješenja je sporija predikcija jer se za jednu predikciju u pozadini obavlja predikcija sa svakim modelom.

Još jedna metoda kojom se postižu bolji rezultati, a koju koriste [14, 17, 18, 19], je TTA (eng. *Test Time Augumentation*) [21]. Ideja ove metode je da se u fazi evaluacije modela na testnom skupu svaka slika, kojoj je potrebno odrediti klasu, izmijeni pomoću jedne ili više transformacija odnosno vrši se augmentacija originalne slike. Konačna klasifikacija originalne slike je srednja vrijednost klasifikacija originalne te klasifikacija transformiranih slika.

Tablica 2.2. prikazuje točnosti dostupnih rješenja, koja su navedena u samim radovima, na FER-2013 skupu podataka. Neka od tih rješenja reproducirana su u sklopu ovoga rada, detaljniji opis toga nalazi se u sljedećem poglavlju.

Tab. 2.2. Usporedba točnosti, na FER-2013 skupu podataka, navedena u pojedinom radu

Metoda	**Točnost**
[14] Kombinacija modela	76.82 %
[15]	75.80 %
[20]	75.42 %
[16]	75.20 %
[17]	74.42 %
[14] Samostalan model	74.14 %
[19]	73.70 %
[18]	73.28 %
[13]	71.16 %

3. OPIS PRAKTIČNOG DIJELA RADA

U ovom poglavlju opisan je način evaluacije rješenja, postupak reprodukcije, uočene prednosti i mane pojedinih rješenja, usporedba konačnih rezultata te pregled aplikacije koja koristi jedno od reproduciranih rješenja.

Odabir rješenja koja su reproducirana u ovome radu ovisio je o uspjehu rješenja. Naime cilj je bio reproducirati što uspješnija rješenja iz područja ovoga rada. Nadalje, odabir je ovisio i o dostupnosti programskog koda pojedinog rješenja koji je trebao biti napisan u Python programskom jeziku budući da je bilo potrebno također u ovome radu izraditi aplikaciju za prepoznavanje emocija na temelju slike ljudskog lica u Python programskom jeziku. Prednost je također dana onim modelima za koje su dostupne trenirane težine. Navedeni uvjeti odabira rješenja su sa ciljem što vjernijeg reproduciranja rješenja kako bi se čitatelju ovoga rada dao pravi uvid u svako rješenje te težinu njegove reprodukcije ukoliko je reprodukcija moguća, te kako bi konačna aplikacija bila što uspješnija.

3.1. Opis evaluacije rješenja

Evaluacija rješenja vrši se na testnom skupu FER-2013 skupa podataka [1] koji u originalnoj verziji skupa podataka nosi naziv „*PrivateTest*“. Metrike koje se koristiti za evaluaciju nalaze se u sljedeća tri potpoglavlja.

3.1.1. Točnost

Točnost (eng. *accuracy*) je definirana sljedećim izrazom:

$$točnost = \frac{TP}{ukupan\ broj\ klasificiranih\ slika} \tag{3-1}$$

gdje TP predstavlja točno klasificirane slike tj. slike klasificirane emocijom koja je zaista prisutna na licu koje se nalazi na slici.

3.1.2. Matrica zabune

Matrica zabune (eng. *confusion matrix*) daje dobar pregled uspješnosti predikcije pojedine klase. Ova matrica se sastoji od onoliko stupaca i redaka koliko ima klasa, gdje svaki stupac i redak predstavljaju jednu klasu i to tako da prvi stupac i prvi redak predstavljaju istu klasu, drugi stupac i drugi redak

predstavljaju drugu klasu i tako sve do zadnjeg stupca i retka. Klase u stupcima predstavljaju pretpostavljene klase dok klase u redcima predstavljaju stvarne klase. U ćelijama na glavnoj dijagonali nalazi se postotak ispravnih predikcija za klasu čiji stupac i redak se sijeku u promatranoj ćeliji. U svim ostalim ćelijama nalazi se postotak neispravnih predikcija gdje se lako može saznati koja klasa je stvarna a koja je pretpostavljena tako što se pronađe redak i stupac koji se sijeku u promatranoj ćeliji.

3.1.3. Vrijeme predikcije

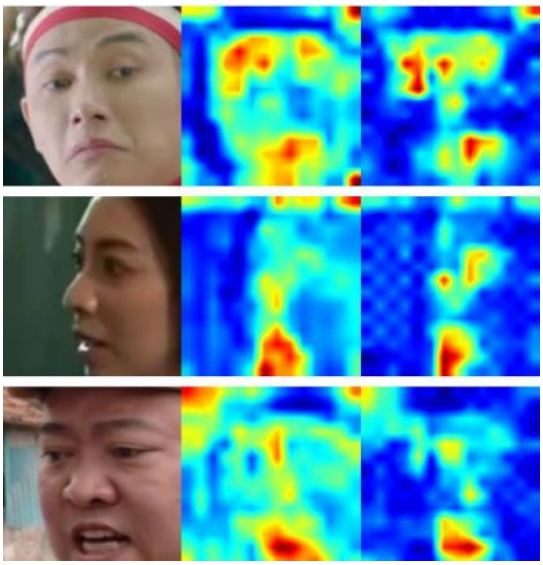
Vrijeme predikcije daje uvid u brzinu predikcije pojedinog modela, vrijeme koje je iskazano za pojedini model je prosječno vrijeme potrebno za predikciju jedne slike, pri tome u to vrijeme nije uračunato vrijeme dodatne obrade podataka koji ulaze u model i podataka koji izlaze iz modela.

Mjerenja su izvršena na računalu sa Ubuntu 22.04.5 LTS operacijskim sustavom koje posjeduje NVIDIA RTX A5000 grafičku karticu, Ryzen Threadripper PRO 3975WX procesor te 128 GB radne memorije.

3.2. Reprodukcija postojećih rješenja

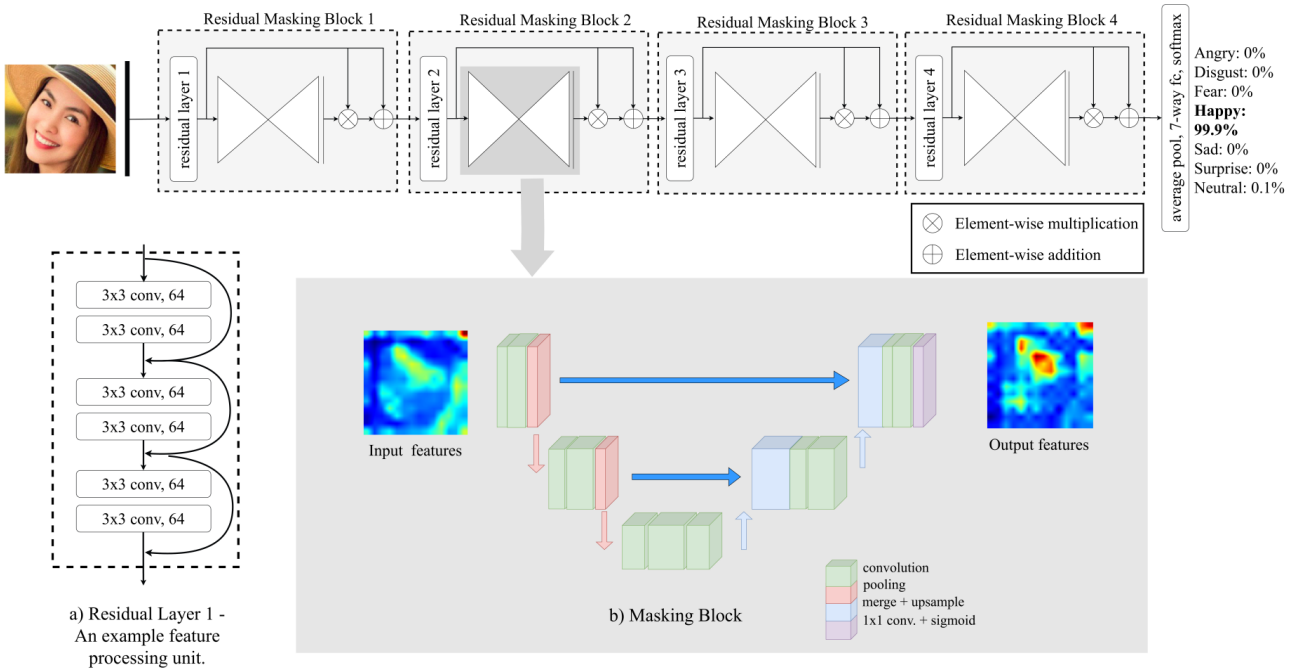
3.2.1. Residual Masking Network

Ovo rješenje opisano je u [14]. Predložena je mreža koji se bazira na kombinaciji U-Net [22] i ResNet34 [23] mreže a polazi od činjenice da za prepoznavanje emocije iz ljudskog lica nisu jednako bitni svi dijelovi lica već su neki dijelovi lica bitniji. Zbog toga ovo rješenje koristi prilagođenu verziju U-Net mreže pomoću koje vrši segmentaciju odnosno mapiranje bitnijih dijelova lica za prepoznavanje emocija. Tako označeni dijelovi lica prikazani su na slici 3.1., dok je arhitektura ove mreže prikazana na slici 3.2.



a) b) c)

Slika 3.1. Usporedba originalne slike i mape značajki a) originalna slika, b) mapa značajki prije 3. Residual Masking bloka, c) mapa značajki poslije 3. Residual Masking bloka [14]



Slika 3.2. Arhitektura Residual Masking mreže [14]

Poveznica na GitHub repozitorij ove metode nalazi se u [14]. U repozitoriju se nalazi i poveznica na pohranu u oblaku gdje je moguće preuzeti prethodno trenirane težine za ovu mrežu. Za reprodukciju ovog rješenja bilo je potrebno klonirati GitHub repozitorij, preuzeti dostupne težine i skup podataka

te napisati vlastitu funkciju za provedbu testiranja koja poziva neke od funkcionalnosti implementiranih u originalnom programskom kodu iz originalnog repozitorija. Repozitorij u kojem je dodana funkcija za provedbu testiranja dostupan je na [24].

Navedena točnost od 74.14 % uspješno je reproducirana no takva točnost dobivena je uz korištenje TTA metode, koja je objašnjena u potpoglavlju 2.3., pa je predikcija svake pojedine slike iz testnog skupa zapravo suma predikcija 10 slika koje su nastale rotacijom i/ili okretanjem originalne slike. Matrica zabune reproduciranog rješenja nalazi se na slici 3.3.

Slika 3.3. Matrica zabune Residual Masking Network rješenja

Reproduciranje ovog rješenja izvršeno je pomoću Google Colaboratory servisa [25] gdje je korištena NVIDIA T4 grafička kartica.

3.2.2. Residual Masking Network u kombinaciji sa ostalim modelima

Kako bi povećali točnost, autori iz [14] predložili su rješenje u kojem je Residual Masking konvolucijska neuronska mreža, koja je opisana u prethodnom potpoglavlju, povezana sa još šest drugih konvolucijskih neuronskih mreža. Tako dobiveno rješenje trenutno postiže najbolje rezultate na FER-2013 testnom skupu podataka sa navedenom točnosti od 76.82 %. Radi se o kombinaciji od četiri ResNet [23], jedne EfficientNet [26] te dvije Residual Masking konvolucijske neuronske mreže od kojih je jedna Residual Masking mreža identična onoj iz prethodnog potpoglavlja dok je druga

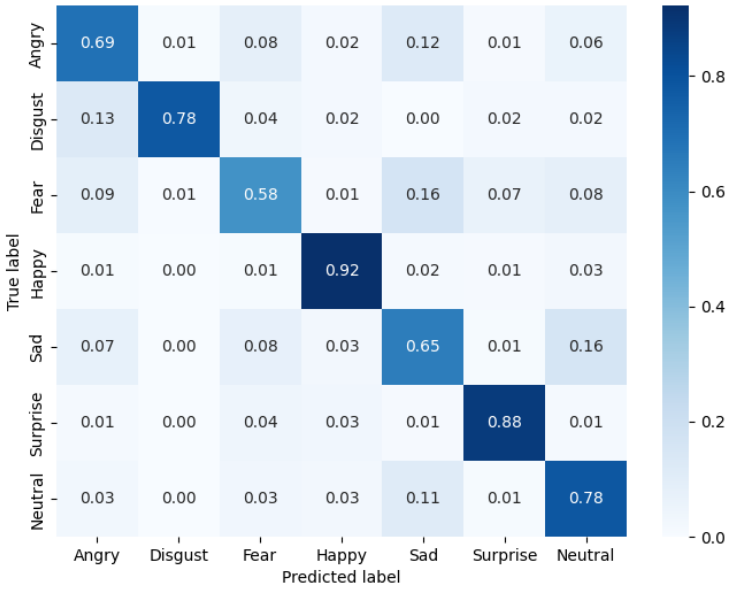
izmijenjena na način da je izostavljen *dropout* sloj. Ovih sedam konvolucijskih neuronskih mreža rade zajedno na način da svaka mreža, za pojedini ulaz, daje vlastiti vektor sa vjerojatnostima pojedine klase. Potom se elementi na istim mjestima ovako dobivenih vektora od svakog modela zbroje što čini konačnu predikciju modela. Ovo rješenje također koristi TTA metodu gdje svaki pojedini model za svaku pojedinu sliku vrši predikciju na osam slika generiranih rotacijom i/ili okretanjem originalne slike. Bitno je naglasiti da je implementacija ovog rješenja na autorovom GitHub repozitoriju *off-line* implementacija što bi značilo da se testni skup prvo evaluira sa svakim pojedinim modelom gdje se izlazi modela spremaju i tek nakon što svi modeli daju predikcije za cijeli testni skup generiraju se konačne predikcije, odnosno modeli ne vrše klasifikaciju paralelno.

GitHub repozitorij ovog rješenja isti je kao i od prethodnog rješenja te se može naći u [14]. U navedenom repozitoriju nalazi se poveznica za preuzimanje težina modela koje se koriste u ovome rješenju. Prilikom reproduciranja rješenja ustanovljeno je da težine za dva modela nedostaju, zbog toga je, u sklopu ovog rada, kontaktiran autor ovoga rješenja sa ciljem ustupljivanja nedostajućih težina. Autor je odgovorio kako ne može pronaći zahtijevane težine modela te ih zbog toga ne može ustupiti. Bez navedenih težina koje nedostaju nije bilo moguće identično reproducirati ovo rješenje čak ni sa dostupnim programskim kodom jer jedan od modela, čije težine nedostaju, za početne vrijednosti težina prilikom treniranja koristi prethodno trenirane težine koje autor ovog rješenja također nije mogao ustupiti.

Bez obzira na činjenicu da identična reprodukcija rješenja nije moguća, reproduciranje rješenja je nastavljeno kako bi se utvrdilo koja je točnost ovog modela na FER-2013 testnom skupu sa dostupnim programskim kodom i težinama. Modeli kojima nedostaju težine trenirani su na Google Colaboratory servisu [25] uz to da su težine modela, koji je u originalnom rješenju za početne vrijednosti težina koristio već prethodno trenirane težine, nasumično postavljene. Tijekom reproduciranja rješenja bilo je potrebe za izmjenama programskog koda kako bi se programski kod mogao uspješno izvršiti, izmjene nisu učinjene na samim modelima i ne utječu na točnost modela, tako izmijenjeni programski kod dostupan je na [24].

Reproducirano rješenje postiglo je točnost od 76.07 % što je očekivano manje od točnosti koja je navedena u radu, zbog nedostajućih težina. U sklopu ovog rada željela se potom ustanoviti točnost ovog rješenja koje se sastoji samo od konvolucijskih neuronskih mreža čije težine su dostupne. Tako reproducirano rješenje postiglo je točnost od 76.34 % a uz to je i računski prihvatljivije u odnosu na

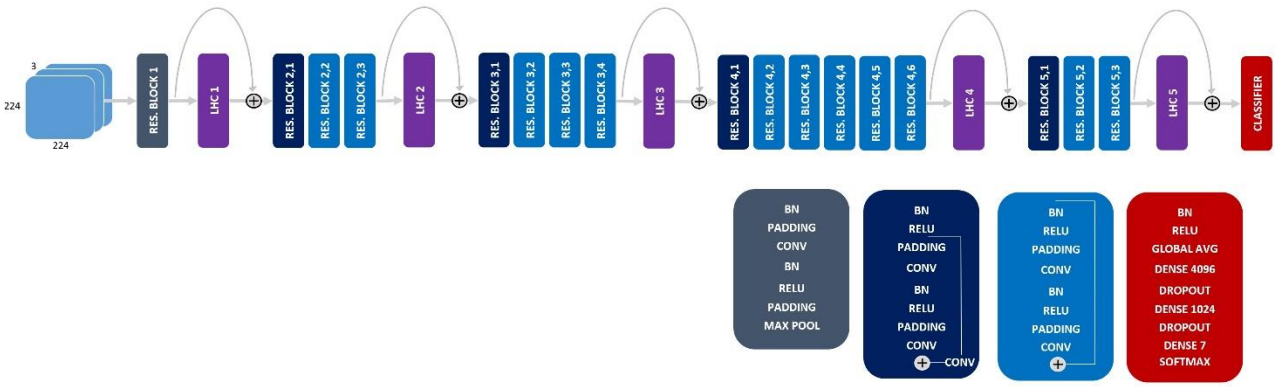
originalno reproducirano rješenje jer se ovo rješenje sastoji od ukupno pet konvolucijskih neuronskih mreža. Ovo rješenje također koristi TTA metodu pa svaki model vrši predikciju na osam slika nastalih rotiranjem i/ili okretanjem originalne slike. Na slici 3.4. prikazana je matrica zabune reproduciranog rješenja koje se sastoji od pet konvolucijskih neuronskih mreža.



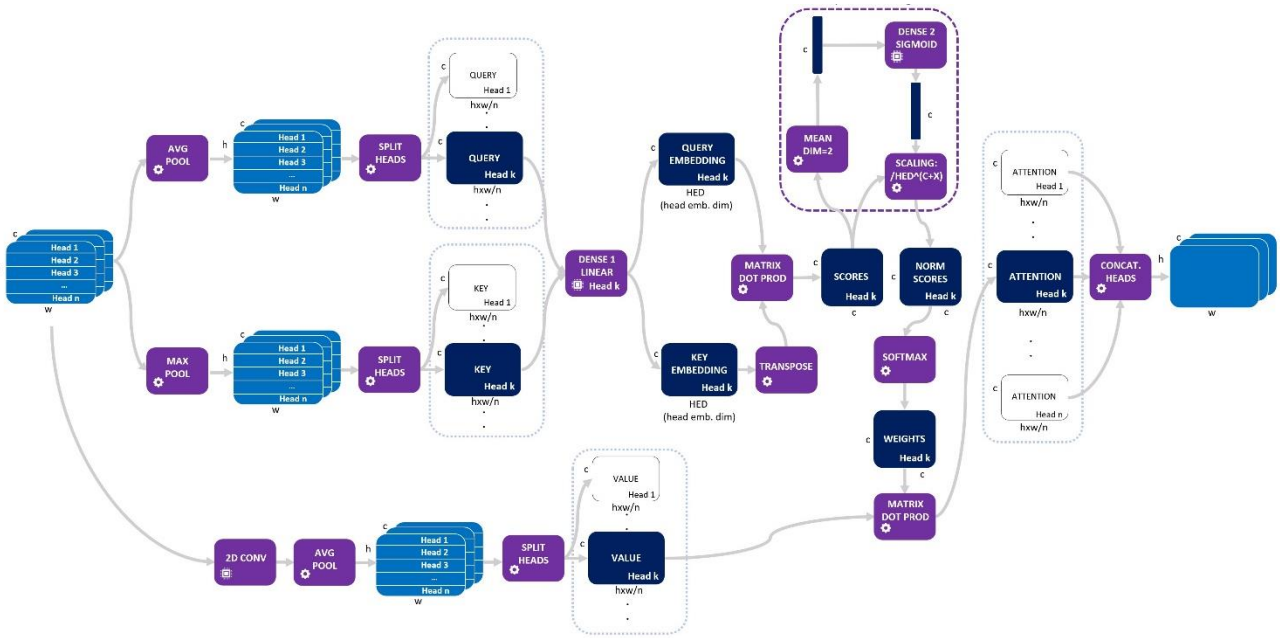
Slika 3.4. Matrica zabune rješenja koje se sastoji od pet konvolucijskih neuronskih mreža

3.2.3. LHC-Net

Rješenje opisano u [17] koristi ResNet34v2 [23] kao glavnu arhitekturu u koju dodaje vlastite module koji se temelje na samopažnji (eng. *self-attention*) [27]. Modul koji je predstavljen u [17] namijenjen je za dodatno poboljšavanje točnosti već prethodno treniranih neuronskih mreža. Takav modul nazivaju *Local (multi) Head Channel (self-attention)* ili skraćeno LHC dok mrežu koja koristi takav modul nazivaju LHC-Net. Kao što je već spomenuto, ovo rješenje koristi ResNet34v2 arhitekturu mreže koja je trenirana za problem prepoznavanja emocija na temelju slike ljudskog lica te je nakon toga u mrežu dodano pet LHC modula nakon čega je bilo potrebno dodatno treniranje mreže. U [17] je navedeno kako je isti postupak treniranja više puta pokrenut te je svaki put dodavanje LHC modula poboljšalo konačan rezultat modela. Slika 3.5. prikazuje arhitekturu mreže dok je na slici 3.6. prikazan LHC modul.



Slika 3.5. Arhitektura ResNet34v2 mreže u koju su dodani LHC moduli [17]



Slika 3.6. LHC modul [17]

Poveznica na GitHub repozitorij ovog rješenja dostupna je u [17]. Rješenje se vrlo lako može reproducirati slijedeći upute koje se nalaze u repozitoriju. Potrebno je pokrenuti nekoliko Python skripti koje automatski preuzimaju FER-2013 skup podataka te prethodno trenirane težine modela nakon čega slijedi evaluacija. Rješenje je reproducirano pomoću računala sa Ubuntu 22.04.5 LTS operacijskim sustavom, NVIDIA RTX A5000 grafičkom karticom te Ryzen Threadripper PRO 3975WX procesorom. Na kraju evaluacije rezultati su onakvi kakvi su navedeni i u [17], reproducirana je točnost modela koja iznosi 74.42 %. Matrica zabune ovog rješenja prikazana je na

slici 3.7. Ovo rješenje također koristi TTA metodu, proučavanjem programskog koda ovog rješenja može se primijetiti kako za svaku ulaznu sliku model napravi predikciju na temelju iste te dodatnih čak 58 slika koje su nastale nekom od transformacija originalne slike što značajno usporava konačnu predikciju.

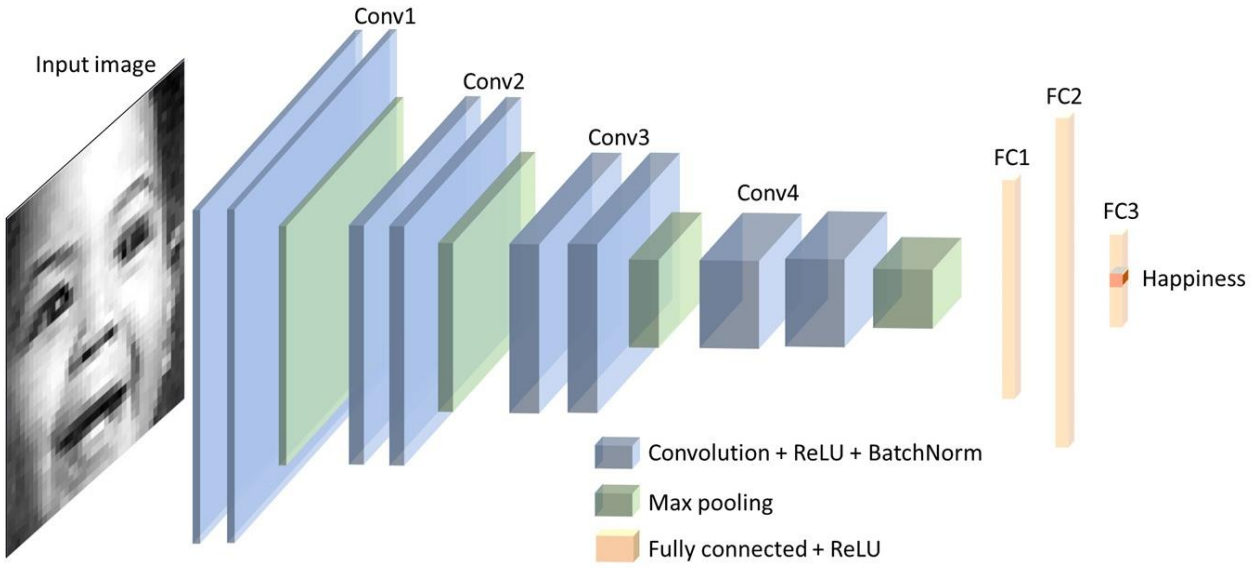
Slika 3.7. Matrica zabune LHC-Net rješenja

3.2.4. VGGNet

Kao što i samo ime ovog potpoglavlja kaže, riječ je o rješenju koje koristi VGG arhitekturu mreže [28]. Ovo rješenje ne predlaže neke novosti po pitanju arhitekture mreže no provodi i opisuje traženje optimalnih hiperparametara, optimalnog algoritma učenja te planera promjene stope učenja sa ciljem postizanja što boljih rezultata za problem prepoznavanja emocija na temelju slike ljudskog lica. Cijeli postupak opisan je u [18]. Konačan model koji je nastao treniranjem sa optimalnim hiperparametrima kao i najboljim algoritmom učenja i planerom promjene stope učenja, kako je navedeno u [18], ostvaruje točnost od 73.06 %. Autorima ovog rješenje to nije bilo dovoljno pa su odlučili dodatno poboljšati rezultat na način da su spojili skup za treniranje i skup za validaciju te provodili treniranje modela na tako proširenom skupu. U [18] se nigdje ne spominje na kojem je skupu, tijekom treniranja na proširenom skupu, vršena validacija, pa jedino ostaje za pretpostaviti da je validacija vršena na testnom skupu. Ovakav način treniranja nije usporediv sa ostalim rješenjima pa se ovaj dio ovog rješenja kao i njegov konačan rezultat neće uzeti u obzir u ovome radu. Dodatna zbunjenost javlja se

kada je u [18] naveden konačan rezultat, za koji se navodi da postiže na FER-2013 testnom skupu podataka, dok je dana matrica zabune za isti model ispod koje piše da je to rezultat postignut na „*PublicTest*“ skupu iz FER-2013 skupa podataka što je zapravo validacijski skup podataka a ne skup za testiranje. Kako je programski kod ovog rješenja također javno dostupan [29] a u opseg ovoga rada spada i reproduciranje rješenja, prilikom pregleda programskog koda primijećeno je kako je u dijelu programskog koda za učitavanje skupa podataka zamijenjen validacijski i testni skup podataka. Ovo saznanje stavlja u pitanje navedene rezultate u spomenutom radu, ne samo u drugoj fazi gdje je za treniranje korišten skup za treniranje i skup za validaciju već i prvi rezultat koji je dobiven uz prethodno pronađene najbolje parametre i algoritam učenja.

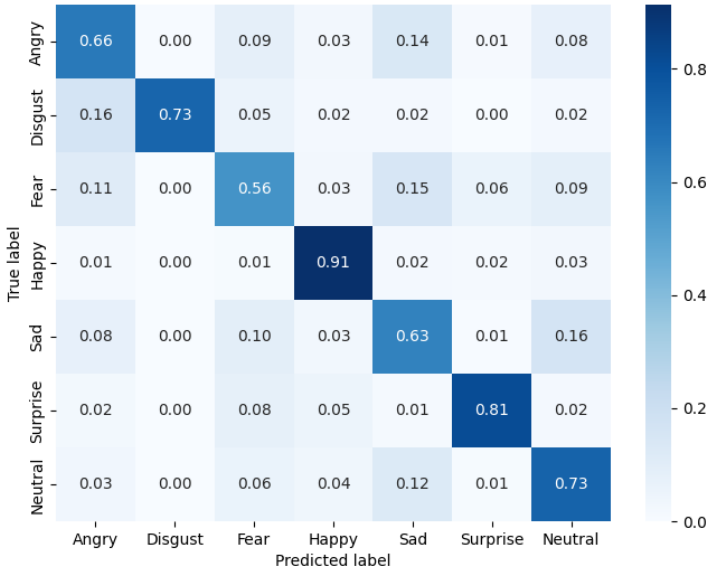
U sklopu ovog diplomskog rada provedeno je treniranje sa dostupnim programskim kodom te sa identičnim parametrima, algoritmom učenja i planera stope učenja kao što je opisano u [18] ali uz prethodnu izmjenu programskog koda kako bi skup za validaciju modela bio zaista validacijski skup FER-2013 skupa podataka, te kako se za validaciju ne bi koristio testni skup podataka kao što se pretpostavlja da je greškom učinjeno u [18]. Arhitektura mreže ovoga rješenja prikazana je na slici 3.8.



Slika 3.8. Arhitektura VGGNet rješenja [18]

Za treniranje je korišteno računalo sa Ubuntu 22.04.5 LTS operacijskim sustavom, NVIDIA RTX A5000 grafičkom karticom te Ryzen Threadripper PRO 3975WX procesorom. Rezultat treniranja sa ispravnim skupovima podataka rezultirao je nešto većom točnosti od one navedene u [18] gdje su

zamijenjeni skupovi podataka za validaciju i testiranje. Točnost koju je ostvario trenirani model iznosi 73.45 %. Ovaj rezultat se odnosi na treniranje u kojemu se ne spajaju skupovi za treniranje i validaciju, već je treniranje provedeno samo na skupu za treniranje u trajanju od 300 epoha. Navedena točnost postignuta je korištenjem TTA metode gdje se odluka o klasi pojedine slike donosi na temelju 10 slika koje su nastale transformacijom originalne slike pomoću TenCrop funkcije [30]. Matrica zabune ovog rješenja prikazana je na slici 3.9.



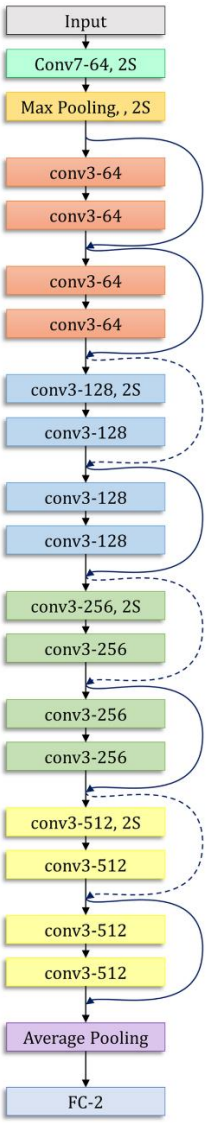
Slika 3.9. Matrica zabune reproduciranog rješenja iz [18]

3.2.5. ResNet18

Zadnje rješenje koje je bilo reproducirano u ovome radu je rješenje dostupno samo na GitHub repozitoriju [19]. Iako ne postoji publikacija u kojoj je opisano ovo rješenje, vrijedi ga spomenuti i reproducirati zbog njegovog dobrog rezultata te dostupnosti programskog koda i treniranih težina. Rješenje koristi standardnu ResNet18 arhitekturu mreže [23] koja je prikazana na slici 3.10.

Rješenje se lako može reproducirati pokretanjem skripte za evaluaciju nakon prethodno preuzetog programskog koda i težina. Težine se mogu preuzeti putem poveznice koja je dostupna na GitHub repozitoriju [19]. Evaluacijom se postiže navedena točnost od 73.7 %, no detaljnijim pregledom programskog koda uočena je sličnost sa programskim kodom rješenja iz prethodnog poglavlja, i to sličnost dijela programskog koda koji služi za učitavanje podataka iz skupa podataka gdje se ponovno

skup koji je namijenjen za validaciju učitava na mjesto skupa za testiranje dok se skup koji je namijenjen za testiranje učitava na mjesto skupa za validaciju. Iz tog razloga bilo je potrebno ponoviti postupak treniranja ovog rješenja koristeći ispravne skupove podataka.



Slika 3.10. ResNet18 arhitektura [31]

Prilikom treniranja, originalni programski kod nije mijenjan osim dijela za učitavanje skupa podataka, gdje su zamijenjeni skupovi za validaciju i testiranje. Parametri i postupak treniranja ostali su jednaki kakvi su korišteni i za treniranje originalnog rješenja. Treniranje je provedeno na računalu sa NVIDIA RTX A5000 grafičkom karticom, Ryzen Threadripper PRO 3975WX procesorom i Ubuntu 22.04.5 LTS operacijskim sustavom. Trenirani model ostvario je točnost od 73.92 % na testnom skupu FER-

2013 skupa podataka. Prilikom testiranja, ovo rješenje koristi TTA metodu transformirajući svaku sliku sa ulaza pomoću TenCrop funkcije [30]. Matrica zabune treniranog modela prikazana je na slici 3.11.

Slika 3.11. Matrica zabune ResNet18 rješenja

3.3. Usporedba rezultata

Nakon pregleda reproduciranja pojedinih rješenja, u ovome poglavlju slijedi usporedba njihovih rezultata. Tablica 3.1. prikazuje točnosti reproduciranih rješenja.

Tab. 3.1. Usporedba točnosti reproduciranih rješenja na FER-2013 testnom skupu podataka

Metoda	**Točnost**
Residual Masking Network u kombinaciji s još četiri modela [14]	76.34 %
LHC-Net [17]	74.42 %
Residual Masking Network samostalan model [14]	74.14 %
ResNet18 [19]	73.92 %
VGGNet [18]	73.28 %

Kada se usporede rezultati iz tablice 2.2 i rezultati iz tablice 3.1. može se primijetiti kako su dva od pet rješenja identično reproducirana, odnosno točnost koja je navedena za LHC-Net i Residual Masking Network mogla se reproducirati koristeći dostupne težine i programski kod. ResNet18 i VGGNet rješenja također se mogu reproducirati no reproducirano rješenje nije usporedivo sa ostalim rješenjima zbog zamjene skupova podataka za testiranje i validaciju te zbog treniranja na združenim skupovima za treniranje i validaciju kao što je slučaj kod VGGNet rješenja. Zbog toga su ResNet18 i VGGNet trenirani na način koji ih čini usporedivim sa ostalim rješenjima koja su reproducirana u ovome radu te su u oba slučaja točnosti na FER-2013 testnom skupu podataka nešto veće od onih navedenih od strane autora ovih rješenja. Rješenje koje kombinira šest konvolucijskih mreža sa Residual Masking mrežom na žalost nije moglo biti reproducirano na identičan način zbog izostanka originalnih težina za dvije konvolucijske mreže, treniranje istih bilo je moguće ali jedna od dvije navedene mreže zahtijevala je prethodno trenirane težine koje također nisu bile dostupne pa su mreže trenirane sa nasumično početno postavljenim težinama. Tako reproducirano rješenje nije davalo dovoljno dobre rezultate. Točnost koja je navedena u tablici 3.1 za Residual Masking Network u kombinaciji još četiri modela točnost je za rješenje koje je reproducirano samo sa dostupnim težinama od autora ovog rješenja što znači da su od ukupno sedam neuronskih mreža, koje su originalno korištene u ovome rješenju, korištene samo njih pet jer je na taj način postignuta najbolja reproducirana točnost ovoga rješenja.

Kao što se može vidjeti u tablici 3.1. i tablici 2.2., rezultati postignuti reproduciranjem rješenja u ovome radu nisu promijenili poredak uspješnosti pojedinih rješenja, odnosno poredak koji su pojedina rješenja imala sa originalno navedenim točnostima održao se i sa reproduciranim rezultatima.

Ono što je zajedničko svim rješenjima je to da koriste TTA metodu kojom poboljšavaju svoju točnost. Sva rješenja najbolje klasificiraju emociju sreće čemu je najvjerojatniji uzrok taj što su slike koje pripadaju u klasu emocije sreće najbrojnije u FER-2013 skupu podataka. Sva rješenja također najlošije klasificiraju emociju straha koju najčešće zamjenjuju sa emocijom tuge.

Zanimljivo je da četiri od pet najnovijih dostupnih rješenja, koja su reproducirana u ovome radu, koriste ResNet arhitekturu mreže što na neki način govori o kvaliteti same arhitekture za problem klasifikacije slike kao i o tome da se ova arhitektura nerijetko vrlo uspješno primjenjuje u problemu prepoznavanja emocije na temelju slike ljudskog lica a dobre rezultate daje unatoč zahtjevnosti referentnog FER-2013 skupa podataka.

Tablica 3.2. prikazuje prosječno vrijeme koje je potrebno da model učini predikciju jedne slike, postupak i okruženje u kojem je izvođeno mjerenje opisano je u potpoglavlju 3.1.3.

Tab. 3.2. Prosječno vrijeme predikcije jedne slike

Metoda	**Vrijeme [ms]**
ResNet18 [19]	0.4
Residual Masking Network samostalan model [14]	9
VGGNet [18]	10
Residual Masking Network u kombinaciji još četiri modela [14]	72
LHC-Net [17]	91

Iz tablice 3.2. vidljivo je da je razlika u vremenu predikcije između pojedinih rješenja puno izraženija u odnosu na razlike u točnosti. Očekivano je da rješenja kao što su LHC-Net i Residual Masking Network u kombinaciji sa još četiri modela imaju veće vrijeme predikcije jer kombinacija više modela zahtjeva i više vremena za predikciju dok LHC-Net konačnu predikciju vrši tek nakon što učini predikciju na 59 slika koje su nastale iz slike na ulazu.

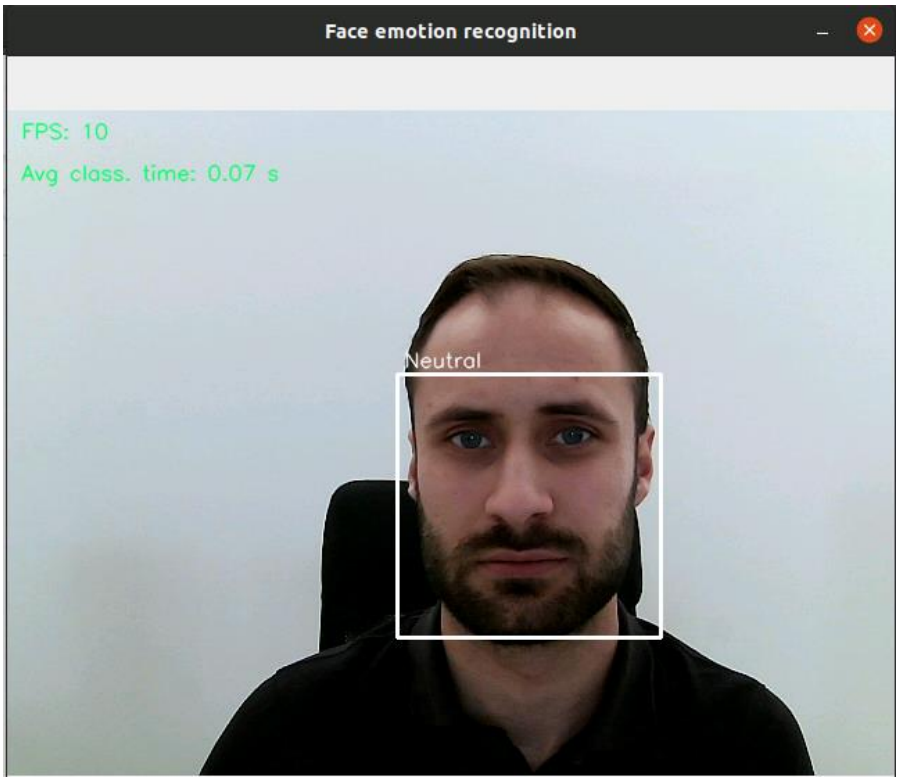
3.4. Pregled razvijene aplikacije za prepoznavanje emocija s ljudskog lica

U ovome poglavlju nalazi se pregled aplikacije za prepoznavanje emocija sa ljudskog lica u stvarnom vremenu koja je nastala kao konačan rezultat ovog rada. Aplikacija je napisana u Python programskom jeziku te za svoj rad koristi OpenCV [32] biblioteku za obradu slike, PyTorch [33] biblioteku za duboko učenje te MediaPipe [34] biblioteku koja se koristi za detekciju lica na slici. Odabir modela za prepoznavanje emocija na temelju slike ljudskog lica ovisio je o rezultatima reprodukcije i evaluacije pojedinih rješenja koja su obrađena u ovome radu, stoga se u aplikaciji koristi rješenje koje kombinira Residual Masking Network sa još četiri neuronske mreže [14] jer ono postiže najbolju točnost na referentnom skupu podataka.

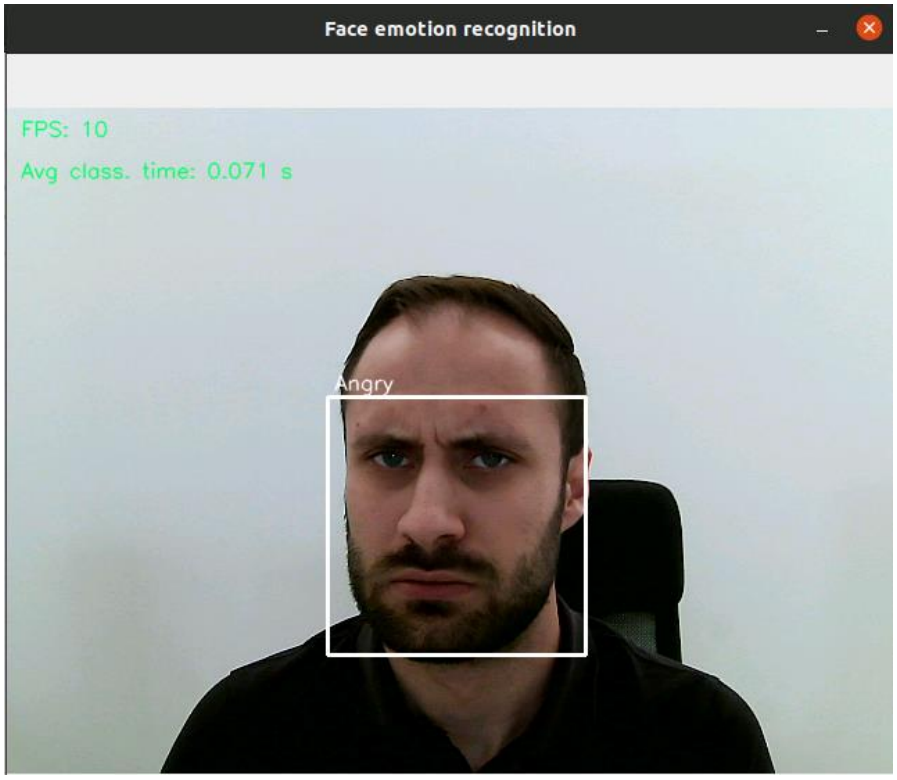
Programski kod aplikacije kao i originalne težine modela nalaze se na DVD-u koji je priložen uz ovaj rad (P.3.1.). Za pokretanje aplikacije potrebno je preuzeti programski kod i težine sa priloženog DVD-

a, kreirati virtualno okruženje pomoću Python alata za kreiranje virtualnog okruženja, preuzeti i instalirati potrebne biblioteke koje se nalaze u *requirements.txt* datoteci te pomoću Python interpretera pokrenuti *realtime_fer.py* datoteku. Preporuča se korištenje Python 3.8 interpretera.

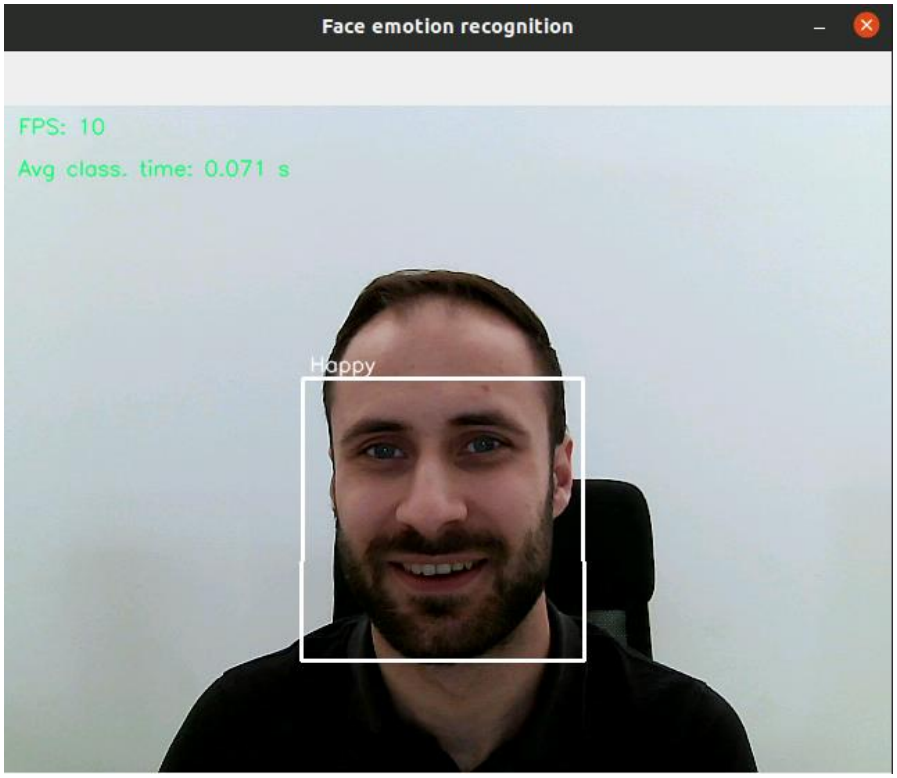
Na slikama 3.12 – 3.14 prikazano je korištenje aplikacije za prepoznavanje emocija sa lica u stvarnom vremenu.



Slika 3.12. Prikaz korištenja aplikacije



Slika 3.13. Prikaz korištenja aplikacije



Slika 3.14. Prikaz korištenja aplikacije

4. ZAKLJUČAK

U sklopu ovoga rada istražena je i opisana problematika prepoznavanja emocija na temelju ljudskog lica. Ova problematika bitna je zbog moguće primjene u interakciji čovjeka i računala gdje bi računalo moglo prilagođavati svoj rad na temelju korisnikova emocionalnog stanja što bi rezultiralo poboljšanjem korisničkog iskustva.

U ovome radu je, za navedeni problem, reproducirano pet dostupnih rješenja koja se temelje na konvolucijskoj neuronskoj mreži. Uspjeh pojedinog rješenja vrednovan je na FER-2013 „PrivateTest“ skupu podataka za koji se slobodno može reći da je referentni skup podataka kada se govori o problemu prepoznavanja emocija na temelju slike ljudskog lica. Reproduciranje rješenja pokazalo se opravdanim jer su samo dva od pet rješenja identično reproducirana, Residual Masking Network kao samostalan model te LHC-Net. Kod VGGNet i ResNet18 rješenja ustanovljeno je da nisu usporediva sa ostalim rješenjima koja su reproducirana u ovome radu zbog zamjene skupa za validaciju i testiranje od strane njihovih autora u procesu nastajanja, stoga su ova rješenja trenirana i evaluirana koristeći ispravne skupove podataka. Rješenje koje koristi Residual Masking Network u kombinaciji sa ostalim modelima reproducirano je sa nešto nižom točnosti od navedene zbog izostanka originalno treniranih težina. Ovo rješenje ostvarilo je najveću točnost no ono je i računalno najzahtjevnije. Na kraju rada izrađena je jednostavna aplikacija koja prepoznaje emocije sa ljudskog lica u stvarnom vremenu a za to koristi rješenje koje kombinira Residual Masking Network i još četiri druge neuronske mreže.

Mjesta za napredak i dalje ima i to u smjeru razvoja efikasnog rješenja koje neće koristiti kombinaciju više mreža te čija će točnost biti slična točnostima reproduciranih rješenja u ovome radu ali bez TTA metode koja usporava konačnu predikciju. Prilikom razvoja ovakvog rješenja bilo bi dobro koristiti ResNet arhitekturu mreže s obzirom da četiri od pet reproduciranih rješenja u ovome radu koriste baš tu arhitekturu.

LITERATURA

[1] Challenges in Representation Learning: Facial Expression Recognition Challenge, Kaggle, https://www.kaggle.com/competitions/challenges-in-representation-learning-facial-expression-recognition-challenge/data, pristupljeno 30.5.2023.

[2] A. Mehrabian, S.R. Ferris, „Inference of attitudes from nonverbal communication in two channels“, J Consult Psychol, Lipanj 1967.

[3] P. Ekman, "Universals and cultural differences in facial expressions of emotion", *Nebraska symposium on motivation*, *19,* 207–283, 1971.

[4] I. J. Goodfellow i suradnici, „Challenges in representation learning: A report on three machine learning contests“, Neural Information Processing: 20th International Conference, 2013.

[5] C. Vinola, K. Vimaladevi, „A Survey on Human Emotion Recognition Approaches, Databases and Applications“, Electronic Letters on Computer Vision and Image Analysis, 2, 24-44, Prosinac 2015.

[6] M. Dubey, L, Singh, „Automatic Emotion Recognition Using Facial Expression: A Review“, International Research Journal of Engineering and Technology, 3, 488-492, Veljača 2016.

[7] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, M. R. Wrobel, „Emotion recognition and its applications“, *Human-Computer Systems Interaction: Backgrounds and Applications 3*, 51-62, 2014.

[8] FER2013 (Facial Expression Recognition 2013 Dataset), Papers with Code, https://paperswithcode.com/dataset/fer2013, pristupljeno 12.6.2023.

[9] A. Mollahosseini, B. Hasani, M. H. Mahoor, „AffectNet: A New Database for Facial Expression, Valence, and Arousal Computation in the Wild”, IEEE Transactions on Affective Computing, 2017.

[10] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, „The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 94–101, 2010.

[11] Image Classification Facial Expression, Challenges in Representation Learning: Facial Expression Recognition Challenge Kaggle, https://www.kaggle.com/code/sharadhaviswanathan/imageclassification-facialexpression, pristupljeno 17.6.2023.

[12] Microsoft/FERPlus, GitHub, https://github.com/Microsoft/FERPlus, pristupljeno 17.6.2023.

[13] Y. Tang, „Deep learning using linear support vector machines“, Workshop on Challenges in Representation Learning, ICML, 2013.

[14] L. Pham, T. H. Vu, T. A. Tran, "Facial Expression Recognition Using Residual Masking Network," 2020 25th International Conference on Pattern Recognition (ICPR), 4513–4519, 2021.

[15] A. Khanzada, C. Bai, F. T. Celepcikay, „Facial expression recognition with deep learning“ arXiv preprint arXiv:2004.11823, 2020.

[16] C. Pramerdorfer, M. Kampel, „Facial expression recognition using convolutional neural networks: state of the art“, arXiv preprint arXiv:1612.02903, 2016.

[17] R. Pecoraro, V. Basile, V. Bono, „Local multi-head channel self-attention for facial expression recognition“, Information, 13(9), 419, 2022.

[18] Y. Khaireddin, Z. Chen, „Facial emotion recognition: State of the art performance on FER2013“, arXiv preprint arXiv:2105.03588, 2021.

[19] LetheSec, Fer2013-Facial-Emotion-Recognition-Pytorch, GitHub, https://github.com/LetheSec/Fer2013-Facial-Emotion-Recognition-Pytorch, pristupljeno 22.6.2023.

[20] M. I. Georgescu, R. T. Ionescu, M. Popescu, „Local learning with deep and handcrafted features for facial expression recognition“, IEEE Access, 7, 64827–64836, 2019.

[21] Test Time Augmentation (TTA) and how to perform it with Keras, Towards Data Science, https://towardsdatascience.com/test-time-augmentation-tta-and-how-to-perform-it-with-keras-4ac19b67fb4d, pristupljeno 23.6.2023.

[22] O. Ronneberger, P. Fischer, T. Brox, „U-net: Convolutional networks for biomedical image segmentation,” International Conference on Medical image computing and computer-assisted intervention, Springer, 234 – 241, 2015.

[23] K. He, X. Zhang, S. Ren, J. Sun, „Deep residual learning for image recognition," proceedings of the IEEE conference on computer vision and pattern recognition, 770–778, 2016.

[24] igorRadic, ResidualMaskingNetwork, GitHub, https://github.com/igorRadic/ResidualMaskingNetwork, pristupljeno 26.6.2023.

[25] Google Colaboratory, https://colab.google, pristupljeno 26.6.2023.

[26] M. Tan, Q. Le, „Efficientnet: Rethinking model scaling for convolutional neural networks", International conference on machine learning, PMLR, 6105–6114, 2019.

[27] A. Vaswani i ostali, „Attention is all you need", Advances in neural information processing systems, 30, 2017.

[28] K. Simonyan, A. Zisserman, „Very deep convolutional networks for large-scale image recognition", 3rd International Conference on Learning Representations, 2015.

[29] usef-kh, fer, Github, https://github.com/usef-kh/fer, pristupljeno 8.7.2023.

[30] TenCrop, Pytorch, https://pytorch.org/vision/master/generated/torchvision.transforms.TenCrop.html, pristupljeno 14.7.2023.

[31] R. Kundu, R. Das, Z. W. Geem , G-T. Han, R. Sarkar, „Pneumonia detection in chest Xray images using an ensemble of deep learning Models", PLoS ONE 16(9), 2021.

[32] OpenCV, https://opencv.org/, pristupljeno 24.7.2023.

[33] PyTorch, https://pytorch.org/, pristupljeno 24.7.2023.

[34] MediaPipe, https://developers.google.com/mediapipe, pristupljeno 24.7.2023

SAŽETAK

Emocionalno stanje čovjeka može se prepoznati iz izraza čovjekova lica. Iako čovjek može doživjeti puno različitih emocija, sve se one mogu svrstati u neke od glavnih kao što su ljutnja, gađenje, strah, radost, tuga, iznenađenost te neutralnost. Problem prepoznavanja emocija sa ljudskog lica u računarstvu moguće je riješiti uporabom strojnog učenja gdje će istrenirani model za sliku čovjekova lica izreći emociju koja prevladava. Ovo područje je zanimljivo zato što može poboljšati interakciju čovjeka i računala. U isto vrijeme je i vrlo izazovno zbog različitih uvjeta u kojima se uzorkuju slike na kojima je potrebno prepoznati emociju, sličnosti između nekih emocija te subjektivnog dojma koji kod ljudi utječe na prepoznavanje emocija. U ovome radu reproducirano je pet postojećih *state-of-the-art* rješenja koja se temelje na konvolucijskoj neuronskoj mreži. Najbolje rješenje korišteno je za konačnu aplikaciju koja prepoznaje emocije sa ljudskog lica u stvarnom vremenu. Uspješnost pojedinih rješenja evaluirana je na FER-2013 skupu podataka.

Ključne riječi: prepoznavanje emocija, CNN, strojno učenje

ABSTRACT

The emotional state of a person can be recognized from the expression on their face. Even though a person can experience a wide range of different emotions, these emotions can be classified into several main classes such as anger, disgust, fear, joy, sadness, surprise, and neutrality. The problem of recognizing emotions from human face in computer science can be solved by using machine learning, where a trained model for facial images can identify the predominant emotion in a person's picture. This field is intriguing because it can enhance human-computer interaction, but at the same time, it is quite challenging due to the various conditions under which images are sampled for emotion recognition, the similarities between some emotions, and the subjective impression that affects human's recognition of emotions. In this master thesis, five existing state-of-the-art solutions based on convolutional neural networks were reproduced, and the best solution was used for the final application that recognizes emotions from human face in real-time. The performance of each solution was evaluated using the FER-2013 dataset.

Keywords: emotion recognition, CNN, machine learning

ŽIVOTOPIS

Igor Radić rođen je u Rijeci 7. svibnja 1998. Osnovnu školu pohađao je u Iloku. 2013. godine upisuje Srednju školu Ilok, smjer tehničar za računarstvo. 2017. godine upisuje preddiplomski sveučilišni studij računarstva na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija. 2020. godine uspješno završava preddiplomski studij te upisuje diplomski sveučilišni studij računarstva modul robotika i umjetna inteligencija.

Potpis:

_____

PRILOZI

- P.3.1. Programski kod aplikacije za prepoznavanje emocija sa ljudskog lica u stvarnom vremenu (priloženo na DVD-u)