

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Metode globokega učenja za
biometrično razpoznavanje na podlagi
očesa

Peter Rot

Ljubljana, 2018

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

**Metode globokega učenja za
biometrično razpoznavanje na podlagi
očesa**

Peter Rot

Delo je pripravljeno v skladu s Pravilnikom o podeljevanju
Prešernovih nagrad študentom

MENTOR: izr. prof. dr. Peter Peer
SOMENTOR: izr. prof. dr. Vitomir Štruc

Ljubljana, 2018

AVTORSKE PRAVICE. Rezultati magistrskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorjev.

©2018 PETER ROT

ZAHVALA

Zahvala pri nastanku tega magistrskega dela gre v prvi vrsti mentorju izr. prof. dr. Petru Peeru in somentorju izr. prof. dr. Vitomirju Štrucu. Hvaležen sem za njuno potrpežljivo usmerjanje pri delu, kvalitetne nasvete, hitro odzivnost ter zagotavljanje vzpodbudnega okolja, v katerem sem se kot študent zelo dobro počutil. Neprecenljive izkušnje pri skupnih projektih bom imel v trajnem, lepem spominu. Naslednja zahvala gre doktorskima študentoma asistentu Žigi Emeršiču, mag., in mlademu raziskovalcu Klemenu Grmu, mag., ki sta mi bila ves čas na voljo in mi pomagala s kreativnimi nasveti in usmeritvami. Sošolcu in prijatelju Dejanu Štepcu, mag., se zahvaljujem, ker je bil na računalniških tekmovanjih požrtvovalen član naše ekipe. Hvaležen sem vsem prostovoljcem, ki so dovolili, da sem fotografiral njihove oči in jih vključil v podatkovno zbirko. Pri obširnem označevanju podatkovne zbirke so veliko dela nase prevzeli študentje Marko Amon, Lovro Jevnikar, Matej Fajdiga, Matej Baša, Lenart Papež in Simon Oblak, za kar se jim iskreno zahvaljujem. Veliko tehničnih znanj, ki sem jih lahko uporabil pri tem delu, sem pridobil v podjetju Philips Digital Pathology v Belfastu. Hvala, ker sem bil lahko del vaše nepozabne ekipe! Katarini Kejžar, prof. slovenskega jezika in mag. govora, se zahvaljujem, ker si je takoj vzela njen dragocen čas in opravila lektoriranje.

Posebno hvaležnost čutim do družinskih članov, ki so mi v študijskih letih trdno stali ob strani. Očetu Hubertu in mami Anamariji se zahvaljujem za brezpogojno ljubezen, vsestransko podporo in toplo domače zavetje. Dekletu Martini iskrena hvala za nesebično pomoč, potrpežljivost in razumevanje, sestri Zali in Tadeju pa za vzpodbudo v težjih trenutkih, ki jih je bilo zaradi njunega smisla za humor moč prebroditi na bolj zabaven način. Zahvaljujem se tudi vsem drugim, ki so kakorkoli prispevali k nastanku tega dela.

Peter Rot, 2018

"Oči so ogledalo duše."

— Latinski pregovor

Povzetek

Naslov: Metode globokega učenja za biometrično razpoznavanje na podlagi očesa

Avtor: Peter Rot

Uspešnost biometričnih sistemov, ki razpoznavajo na podlagi očesnih modalnosti, je kritično odvisna od pogojev zajema slike in od natančnosti postopka segmentacije. Za zmanjšanje napak tovrstnih biometričnih sistemov potrebujemo odporne segmentacijske metode. Šarenica je bila kot biometrična modalnost zaradi visoke razpoznavalne natančnosti v preteklem desetletju deležna velike pozornosti, raziskovalci pa so kot samostojne modalnosti (ali dopolnilne šarenici) predlagali tudi beločnico in periokularno regijo.

Z metodami globokega učenja, ki so se na mnogih področjih računalniškega vida izkazale kot najbolj uspešne, v tem delu obravnavamo vsako modalnost posebej (beločnico, periokularni del, šarenico), nato pa s fuzijo vse tri združimo v enoten razpoznavalni cevovod. Glaven poudarek dela je na razpoznavanju iz beločnice, pri katerem *i*) izdelamo novo podatkovno zbirko SBVPI, *ii*) predlagamo segmentacijske metode, s katerimi smo dvakrat osvojili prvo mesto na tekmovanjih SS(ER)BC, ter *iii*) razvijemo in ovrednotimo preostanek cevovoda za razpoznavanje na podlagi beločnice. Predlagamo metodo za večrazredno segmentacijo očesa, s katero dosežemo vzpodbudne rezultate. Nato predlagamo in ovrednotimo cevovod za razpoznavanje na podlagi periokularnega dela, za razpoznavanje na podlagi šarenice pa uporabimo že obstoječ cevovod. Na koncu ovrednotimo še fuzijo vseh treh modalnosti.

Z metodami globokega učenja dosežemo obetavne razpoznavalne natančnosti za vsako izmed treh modalnosti. Z združevanjem modalnosti v skupen fuzijski sistem pa razpoznavno natančnost dodatno izboljšamo.

Ključne besede

globoko učenje, konvolucijske nevronske mreže, beločnica, šarenica, periokularna informacija, segmentacija, razpoznavanje

Abstract

Title: Deep learning methods for biometric recognition based on eye information

Author: Peter Rot

The accuracy of ocular biometric systems is critically dependent on the image acquisition conditions and segmentation methods. To minimize recognition error robust segmentation algorithms are required. Among all ocular traits, iris got the most attention due to high recognition accuracy. New modalities such as sclera blood vessels and periocular region were also proposed as autonomous (or iris-complementary) modalities.

In this work we tackle ocular segmentation and recognition problems using deep learning methods, which represent state-of-the-art in many computer vision related tasks. We individually evaluate three recognition pipelines based on different ocular modalities (sclera blood vessels, periocular region, iris). The pipelines are then fused into a single biometric system and its performance is evaluated. The main focus is sclera recognition in the scope of which we *i*) create a new dataset named SBVPI, *ii*) propose and evaluate segmentation approaches, which won the first place on SS(ER)BC competitions, and *iii*) develop and evaluate the rest of the sclera-based recognition pipeline. The next contribution of this work is multi-class eye segmentation technique, which gives promising results. We also propose and evaluate deep learning pipeline for periocular recognition. For iris recognition we use an existing pipeline and evaluate it on our dataset.

With deep learning we achieve promising recognition results for each individual modality. We further improve recognition accuracy with multi-modal fusion of all three modalities.

Keywords

deep learning, convolutional neural networks, sclera, iris, periocular information, segmentation, recognition

Kazalo

Povzetek

Abstract

Seznam uporabljenih kratic

1	Uvod	1
1.1	Motivacija	1
1.2	Prispevki	2
1.3	Struktura naloge	2
2	Sorodna dela	5
2.1	Anatomija očesa	5
2.2	Očesna biometrija	6
2.3	Razpoznavanje na podlagi beločnice	8
2.4	Razpoznavanje na podlagi periokularnega dela	9
2.5	Razpoznavanje na podlagi šarenice	10
2.6	Biometrična fuzija očesnih modalnosti	11
3	Nevronske mreže	13
3.1	Osnovni gradniki nevronske mreže	13
3.2	Konvolucijske nevrnske mreže	23
4	Podatkovne zbirke	37
4.1	MASD	37
4.2	UBIRIS.v2	38
4.3	SBVPI	38
4.4	Primerjava podatkovnih zbirk	43

KAZALO

5	Prepoznavanje na podlagi beločnice	45
5.1	Uvod	45
5.2	Metode	46
5.3	Eksperimenti in rezultati	53
5.4	Diskusija	68
6	Večrazredna segmentacija očesa	71
6.1	Uvod	71
6.2	Metodologija	72
6.3	Večrazredna segmentacija in učenje	73
6.4	Rezultati večrazredne segmentacije	75
6.5	Analiza občutljivosti in kvalitativni rezultati	77
6.6	Primerjava z enorazredno segmentacijo	80
6.7	Diskusija	81
7	Prepoznavanje na podlagi periokularnega dela	83
7.1	Uvod	83
7.2	Metode	84
7.3	Učenje modelov in generiranje izhoda	87
7.4	Eksperimenti in rezultati	90
7.5	Diskusija	92
8	Prepoznavanje na podlagi šarenice	93
8.1	Uvod	93
8.2	Metode	94
8.3	Rezultati	95
8.4	Diskusija	96
9	Fuzija modalnosti	99
9.1	Uvod	99
9.2	Metode	99
9.3	Rezultati	101
9.4	Diskusija	103
10	Sklepi in nadaljnje delo	105

Seznam uporabljenih kratic

kratica	angleško	slovensko
ANN	artificial neural network	umetna nevronska mreža
AUC	area under the (ROC) curve	ploščina pod (ROC) krivuljo
AUCMC	area under the normalized CMC	ploščina pod normirano CMC krivuljo
BNN	biological neural network	biološka nevronska mreža
CED	convolutional encoder-decoder	konvolucijski kodirnik-dekodirnik
CLAHE	contrast-limited adaptive histogram equalization	kontrastno omejena adaptivna izenačitev histograma
CMC	cumulative match score curve	kumulativna krivulja ujemanja
CNN	convolutional neural network	konvolucijska nevronska mreža
DCT	discrete cosine transform	diskretna kosinusna transformacija
DNN	deep neural network	globoka nevronska mreža
DPI	dots per inch	število pik na palec
EER	equal error rate	ocena enake napake
FAR	false acceptance rate	standardna ocena napačnega sprejemanja
FCN	fully convolutional networks	celostno-konvolucijske mreže
FFT	fast Fourier transform	hitra Fourierjeva transformacija
FNN	feed-forward neural network	aciklična nevronska mreža
FRGC	Face Recognition Grand Challenge	Veliki izziv razpoznavanja obrazov
FRR	false rejection rate	standardna ocena napačne zavrnitve
GIMP	GNU image manipulation program	GNU program za manipulacijo slik
GNU	GNU is not Unix	GNU ni Unix
HWT	Haar valjčna transformacija	Haar wavelet transform
HOG	histogram of oriented gradients	histogram usmerjenih gradientov
HSV	hue, saturation, value	odtenek, zasičenost, vrednost
ICB	International Conference on Biometrics	mednarodna biometrična konferenca

ILSVRC	ImageNet Large Scale Visual Recognition Competition	ImageNet tekmovanje vizualnega razpoznavanja v velikem obsegu
LBP	local binary pattern	lokalni binarni vzorec
LTU	linear threshold unit	linearna pragovna enota
MASD	Multi-Angle Sclera Dataset	Zbirka beločnic, posnetih iz različnih kotov
MICHE	Mobile Iris Challenge Evaluation	Izziv razpoznavanja šarenice z mobilnimi napravami
MLP	multi-layer perceptron	večslojni perceptron
MOS	max-out strategy	strategija izbire maksimalnega elementa
MTL	multi-tasking learning	večkriterijsko učenje
NIR	near-infrared	blizu infrardeče
RCU	residual convolution unit	konvolucijska enota preostankov
ReLU	rectified linear unit	popravljen linearna enota
RGB	red, green, blue	rdeča, zelena, modra
RNN	recursive neural network	rekurzivna nevronska mreža
ROC	receiver operating characteristic	karakteristika delovanja sprejemnika
ReLU	rectified linear unit	popravljen linearna enota
SGD	stochastic gradient descent	stohastični gradientni spust
SBVPI	Sclera Blood Vessels, Periocular and Iris dataset	Zbirka žil iz beločnice, periokularnih regij in šarenic
SIFT	scale-invariant feature transform	na skaliranje invariantna transformacija značilk
SS(ER)BC	Sclera Segmentation (and Eye Recognition) Benchmarking Competition	Tekmovanje segmentacije beločnice (in razpoznavanja na podlagi očesa) s primerjalnimi analizami
TAR	true acceptance rate	standardna ocena pravega sprejemanja
THS	thresholding strategy	strategija upragovanja
USS	unsupervised sclera segmentation CMC	nenadzorovana segmentacija beločnice
VER	verification	standardna verifikacijska ocena

Poglavje 1

Uvod

1.1. Motivacija

Med biometričnimi modalnostmi (recimo prstni odtis, obraz, mrežnica, način hoje) šarenica po mnogih kriterijih [1] velja za eno izmed najbolj primernih modalnosti, saj na njeni podlagi lahko z računalniškimi sistemi zelo natančno določimo identiteto osebe, če je sistem šarenico te osebe videl že v preteklosti. Prepoznavanje na podlagi šarenice (angl. iris recognition) zato postaja vse bolj razširjen način za identifikacijo ter avtentikacijo. Stabilnost in uporabnost tovrstnih sistemov je kritično odvisna od pogojev, v katerih sliko zajamemo, ter od kvalitete procesa segmentacije. Cilj segmentacije je za posamezen slikovni element na sliki celotnega očesa napovedati, kateremu delu očesa pripada – ali prikazuje šarenico, beločnico ali periokularni del. Glavni vir težav povzročajo spremenljivi pogoji pri zajemu slike (oko je bolj ali manj odprto, odboji svetlobe, kamera je glede na oko drugače pozicionirana, oko je delno zakrito ipd.), kar povzroči manjšo natančnost biometričnih sistemov.

V tem delu se bomo zato najprej osredotočili na izdelavo odpornih algoritmov za segmentacijo očesa, ki nam bodo sliko očesa razdelili na več semantičnih delov. Zanimalo nas bo, kakšno natančnost razpoznavanja je možno dobiti le z uporabo dela slike, ki pripada izključno beločnici. Dobljeno natančnost bomo primerjali z natančnostjo razpoznavanja, ki jo dobimo le z uporabo dela slike, ki pripada izključno periokularni regiji. Na podoben način bomo naredili primerjavo z natančnostjo, ki jo pridobimo izključno z uporabo dela slike, ki pripada šarenici. Izdelali bomo tri ločene razpoznavalne cevovode, pri čemer bo vsak cevovod deloval le na podlagi posamezne modalnosti, ter med njimi naredili primerjavo. Vsi cevovodi bodo delovali na metodah globokega učenja. Na ta način bomo

ugotovili, koliko diskriminativne informacije (ki je za razpoznavanje ključnega pomena) smo zmožni izluščiti iz posameznih delov. Ker različni deli očesa prispevajo drugačno diskriminativno informacijo, nas bo v nadaljevanju zanimalo, kako združevanje informacij iz posameznih modalnosti vpliva na končno uspešnost razpoznavanja. Tovrstno združevanje biometričnih modalnosti imenujemo fuzija.

1.2. Prispevki

Prvi prispevek magistrske naloge je nova, javno dostopna podatkovna zbirka, namenjena predvsem raziskovalcem, ki se ukvarjajo z biometričnim razpoznavanjem na podlagi beločnice in periokularne informacije. Ker je na slikah vidna tudi regija šarenice, je zbirka primerna tudi za razpoznavanje na podlagi šarenice.

Drugi prispevek je primerjava zmogljivosti različnih najnovejših segmentacijskih arhitektur (ki bazirajo na metodah globokega učenja) na ožji domeni biometrije očesa. Tovrstnih osredotočenih primerjav zaenkrat še nismo zasledili.

Tretji prispevek je kvalitetna segmentacija delov očesa, torej šarenice, beločnice in periokularne regije, ki se jih neposredno da uporabiti v kasnejših fazah cevovoda za razpoznavanje. Predstavljen in ovrednoten je tudi način za večrazredno segmentacijo očesa. Prispevki naloge imajo potencial povečati odpornost že obstoječih sistemov za razpoznavanje, saj je eden izmed ciljev narediti segmentacijski algoritem čimbolj neobčutljiv na tipične spremembe pogojev pri zajemanju slike, kot so spremembe svetlobe in pozicije osebe glede na kamero.

Četrti prispevek je nov pristop za razpoznavanje na podlagi vaskularnih struktur iz beločnice, ki je v celoti osnovan na metodah globokega učenja.

Peti prispevek je primerjava dveh algoritmov za razpoznavanje na podlagi periokularnega dela, ki delujeta po principih globokega učenja.

Šesti prispevek je uporaba treh ločenih biometričnih modalnosti (šarenice, beločnice in periokularne informacije) v enotnem fuzijskem sistemu za razpoznavo.

1.3. Struktura naloge

V poglavju 2. *Sorodna dela* je najprej predstavljena anatomija očesa, nato pa je narejen širši pregled področja očesne biometrije. Na kratko so predstavljeni anatomske deli očesa, njihova struktura ter osnovni principi njihovega delovanja. Poznavanje le-teh je pomembno za razumevanje nadaljnjega dela. Nato so naštet in na kratko opisane pomembnejša

znanstvena dela, ki se tematsko skladajo s to nalogo, naše delo pa je umeščeno v širši kontekst očesne biometrije. Cilj naše naloge je posebej obravnavati vsako izmed treh očesnih modalnosti, in sicer: *i*) beločnico, *ii*) periokularno regijo ter *iii*) šarenico, nato pa vse tri ločene sisteme združiti v enoten razpoznavalni sistem. Ti cilji tudi določajo nadaljnjo razdelitev poglavja 2. *Sorodna dela* na podpoglavja.

Sledi poglavje 3. *Nevronske mreže*, kjer so opisane teoretične osnove za razumevanje metod globokega učenja. Iz osnovnih gradnikov (kot je npr. nevron) so postopoma izpeljane klasične nevronske mreže. Razloženo je delovanje konvolucijskih nevronskih mrež in predstavljenih je nekaj sodobnih arhitektur, ki temeljijo na le-teh. Iz konvolucijskih nevronskih mrež so naprej izpeljane celostno-konvolucijske mreže, ki jih v naših eksperimentih uporabljamo za segmentacijo.

V poglavju 4. *Podatkovne zbirke* so predstavljene za naše delo pomembne podatkovne zbirke. Pri vsaki obstoječi podatkovni zbirki analiziramo njene prednosti in slabosti, slednje pa so motivacija za izdelavo lastne podatkovne zbirke, s katero smo želeli nasloviti pomanjkljivosti obstoječih zbirk. V tem poglavju natančno opišemo novo podatkovno zbirko SBVPI, ki je primarno namenjena razpoznavanju na podlagi beločnice.

Sledi pet poglavij, ki opisujejo naše metode, eksperimente in rezultate, njihovo zaporedje pa se ujema s prej navedenimi cilji naloge. V poglavju 5. *Prepoznavanje na podlagi beločnice* najprej ovrednotimo različne segmentacijske pristope, ki temeljijo na metodah globokega učenja. Primerjamo jih tudi s segmentacijskim algoritmom, ki temelji na metodah tradicionalnega strojnega učenja. Nato z metodami globokega učenja razvijemo celoten cevovod za razpoznavanje na podlagi beločnice in ga ovrednotimo v primerjavi z algoritmom za razpoznavanje, ki temelji na metodah tradicionalnega strojnega učenja. V poglavju 6. *Večrazredna segmentacija očesa* predlagamo in ovrednotimo pristop za večrazredno segmentacijo očesa, ki slikovne elemente na sliki očesa klasificira v 6 različnih razredov. Zmožljivosti postopka analiziramo glede na različne smeri pogleda, narejena pa je tudi primerjava večrazredne segmentacije z enorazredno segmentacijo, ki se danes tipično uporablja v večini tovrstnih biometričnih cevovodov. V poglavju 7. *Prepoznavanje na podlagi periokularnega dela* primerjamo dva postopka za razpoznavanje na podlagi periokularnega dela, ki sta osnovana na metodah globokega učenja. V poglavju 8. *Prepoznavanje na podlagi šarenice* na kratko opišemo pristop, ki smo ga uporabili za pridobitev značilnk iz predela šarenice. V poglavju 9. *Fuzija modalnosti* združimo vse tri biometrične sisteme na nivoju njihovih ocen (angl. score level fusion) in poročamo o izboljšanju razpoznavanja, ki ga pridobimo s fuzijo.

Sklepi in ideje za nadaljnje delo so podani v poglavju 10. *Sklepi in nadaljnje delo*.

Poglavje 2

Sorodna dela

V tem poglavju je narejen pregled del, ki so najbolj pomembna za naš razpoznavalni cevod. Cilj tega poglavja je predstaviti anatomsko zgradbo očesa, umestiti naše prispevke v širši kontekst področja in predstaviti motivacijo za naše delo.

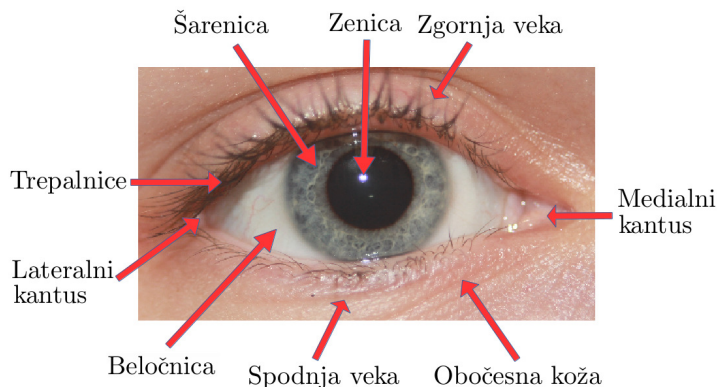
2.1. Anatomija očesa

Na sliki 2.1 so prikazani deli očesa, ki jih vidimo od zunaj. Viden je *i*) del očesnega zrkla (ki ga sestavljajo zenica, šarenica in beločnica) ter *ii*) obočesni oz. periokularni del, pod katerega (s stališča biometrije) spadajo: zgornja in spodnja veka s trepalnicami, medialni in lateralni kantus, obočesna koža, poleg pa lahko prištevamo tudi obrvi.

Celotno očesno zrklo, ki je prikazano na sliki 2.2, lahko v grobem razdelimo na tri dele [2, 3]: vezivni sloj, vaskularni sloj in notranji sloj. Deli si v omenjenem zaporedju sledijo, če zrklo gledamo iz njegovega zunanjšega dela proti notranjosti. Del vlaken, ki predstavlja vezivni sloj, je sestavljen iz dveh zveznih delov: beločnice (angl. sclera) in roženice (angl. cornea). Beločnica je sestavljena iz približno 85% vezivnega tkiva, zato je relativno trdna, kar omogoča, da varuje globlje sloje zrkla. Beločnica je pripeta na mišice izven zrkla (angl. extraocular muscles), ki omogočajo obračanje zrkla. Skozi beločnico prehaja mnogo žil in živcev (npr. optični živec).

Vaskularni sloj zrkla je sestavljen iz treh zveznih delov: žilnice (angl. choroid), ciliarnega telesa (angl. ciliary body) in šarenice (angl. iris). Žilnica je sestavljena iz žil in vezivnega tkiva. Njena glavna naloga je oskrba zunanjih slojev mrežnice (angl. retine). Ciliarno telo med drugim sestavljajo ciliarne mišice, ki omogočajo spreminjanje oblike leče. Ob gledanju na daleč so mišice ciliarnega telesa sproščene (leča je manj konkavna), ob gle-

danju na blizu pa se mišice skrčijo (leča je bolj konkavna). V sredini šarenice je luknja, ki jo imenujemo zenica (angl. pupil). Z večanjem in manjšanjem zenice oko uravnava količino svetlobe, ki lahko pride v oko.



Slika 2.1: Zunanji deli očesa.

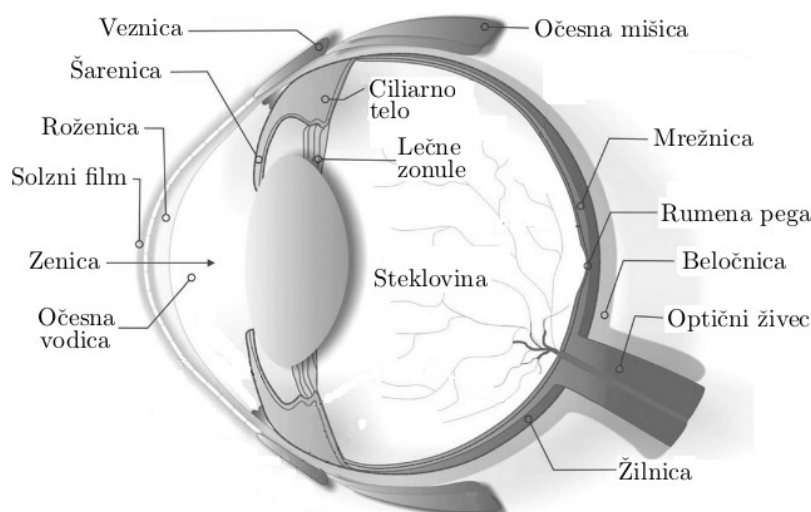
Notranji sloj očesa predstavlja mrežnica (angl. retina), katere naloga je detekcija svetlobe. Mrežnica vsebuje dve vrsti fotoreceptorjev: *i*) paličice (angl. rods) in *ii*) čepke (angl. cones). Paličice so občutljive na vidno svetlobo katerekoli valovne dolžine, čepki pa nam omogočajo, da vidimo barve. Čepki, ki jih je približno 20-krat manj kot paličic, se večinoma nahajajo okoli predela rumene pege. Paličice najdemo tudi na obrobju mrežnice.

Na površju beločnice lahko opazujemo različne tipe žil. Najbolj na površini so veznične žile (angl. conjunctival blood vessels). Pod njimi so vidne nadbeločnične (episkleralne) žile, ki potekajo radialno proti roženici. Najgloblje žile, ki so še vedno vidne, pa ležijo v beločnici.

S staranjem kolagen in elastična vlakna začnejo razpadati ter beločnica začne dehidrirati, struktura žil pa se kljub temu ohrani [4].

2.2. Očesna biometrija

V zadnjem desetletju je področje očesne biometrije zelo napredovalo, večinoma po zaslugi pionirskih del Johna Daugmana na temo razpoznavanja na podlagi šarenice [5, 6, 7]. Predlagal je opisnik značilk (angl. feature descriptor) imenovan IrisCode, ki omogoča visoko natančnost razpoznavanja in hitre izračune ujemanja med identitetami. Kasneje so bili za predstavitev značilk v šarenici predlagani tudi drugi opisniki značilk [8, 9, 10, 11]. V primerjavi z drugimi biometričnimi modalnostmi (npr. prstni odtis, obraz, govor, način hoje) ima šarenica poleg visoke natančnosti razpoznavanja še mnogo drugih prednosti.



Slika 2.2: Anatomija očesnega zrkla [2].

Omogoča brezstičen in neinvaziven zajem informacije, v procesu staranja se relativno malo spreminja, uporabnikom pa se v precejšnji meri zdi sprejemljiva modalnost za zajem [12]. Algoritmi, ki izkoriščajo informacijo šarenice, v večini primerov zagotavljajo visoko natančnost le v primeru, da je slika očesa zajeta v dovolj visoki kvaliteti in ločljivosti [13, 14]. Obstaja še veliko odprtih raziskovalnih vprašanj, ki zadevajo procesiranje slik, ki so zajete po manj omejenih protokolih zajema (npr. oseba ne gleda naravnost v kamero, oči so zaprte, trepalnice zakrivajo regijo šarenice itd.). Da bi zagotovili odpornost tovrstnih sistemov, lahko izrabimo tudi druge očesne modalnosti.

Avtomatski razpoznavalni sistemi, ki razpoznavajo na podlagi drugih očesnih modalnosti (npr. beločnica ali periokularni del), so bili v zadnjih letih deležni velike pozornosti s strani raziskovalcev [12, 15, 16, 17, 18, 19, 20, 21, 22]. Naslednja očesna modalnost, ki omogoča izjemno razpoznavno natančnost, je struktura žil na mrežnici [23]. Ta modalnost se od drugih očesnih modalnosti razlikuje po tem, da je zajem bolj zahteven, saj zanj potrebujemo poseben skener in višjo stopnjo kooperacije subjektov. Poleg tega je manj primerna za širšo uporabo, saj se iz debeline žil da sklepati tudi na zdravstveno stanje osebe (npr. prisotnost diabetesa [24]), kar je v mnogih scenarijih prevelik poseg v zasebnost. Zanimanje za raziskovanje očesnih modalnosti je v porastu predvsem zaradi ogromnega tržnega potenciala očesne biometrije. Veliko del [25, 26, 27, 28] skuša (ob)očesno biometrijo uporabljati za avtentikacijske namene pri uporabi mobilnih naprav.

Konvolucijske nevronske mreže so se v zadnjih letih na mnogih področjih računalniškega vida izkazale kot učinkoviti napovedni modeli, s katerimi je mogoče dosegati najboljše rezultate [29, 30, 31]. Ena izmed ključnih prednosti globokega učenja je, da se mreža

diskriminativnih značilnk lahko iz podanih podatkov nauči tako rekoč sama. To v praksi pogosto pomeni, da za učenje dovolj natančnih modelov potrebujemo veliko označenih podatkov. Danes mnogo raziskovalcev skuša reševati probleme očesne biometrije z uporabo globokega učenja [16, 17, 32, 33, 34, 35].

2.3. Razpoznavanje na podlagi beločnice

2.3.1. Segmentacija beločnice

Literatura, ki opisuje segmentacijo beločnice, je v grobem razdeljena na *i*) tehnike, ki se ukvarjajo izključno s pridobitvijo vaskularne strukture iz vhodnih slik in *ii*) metode, ki poskušajo segmentirati celotno regijo beločnice. Pregled tehnik iz prve skupine je podrobneje opisan v delu [36]. Nas bodo v tem delu zanimale predvsem metode iz druge skupine. Problem segmentacije beločnice je tipično proučevan kot podproblem drugih nalog [37], kot sta na primer prepoznavanje na podlagi šarenice in ocenjevanje smeri pogleda (angl. gaze estimation). Beločnica je lahko uporabljena tudi kot samostojna biometrična modalnost [38] ali kot dopolnilna modalnost šarenici [39]. V začetnih delih (npr. [40, 41]) je bila segmentacija beločnice narejena ročno. V želji, da bi vzpodbudili razvoj avtomatskih segmentacijskih metod in izdelali enoten protokol za primerjalne analize tovrstnih algoritmov, so bila v sklopu najpomembnejših biometričnih konferenc organizirana tekmovanja SS(ER)BC [20, 42, 43]. Večina algoritmov, ki so bili predstavljeni v sklopu teh tekmovanj, temelji na predpostavkah o relacijah med vrednostmi slikovnih elementov za opis rdeče, zelene in modre barve (angl. red, green, blue, RGB) ali med vrednostmi za opis odtenka, zasičenosti in vrednost (angl. hue, saturation, value, HSV). Temeljijo tudi na gručenju (angl. clustering) ter morfoloških operacijah. Tovrstni pristopi so precej neodporni, saj imamo ljudje veliko različnih odtenkov kože, kar na preprost način težko definiramo z relacijami RGB/HSV vrednosti.

Na zadnjih dveh SS(ER)BC tekmovanjih so se kot najbolj učinkovite metode izkazale tehnike, ki temeljijo na konvolucijski arhitekturi kodirnika-dekodirnika (angl. convolutional encoder-decoder, CED). Le-te so bile razvite v okviru našega dela. Tovrstne metode zagotavljajo vrhunske rezultate in od konkurenčnih algoritmov odstopajo z veliko razliko, kar je podrobneje opisano v delih [20, 44]. V sklopu SS(ER)BC tekmovanj je bila uporabljena podatkovna zbirka MASD. Prednost zbirke MASD je predvsem način zajema slik, pri katerem za vsako identiteto zberejo več primerkov slik, ki pripadajo štirim različnim smerem pogleda: pogled naravnost, levo, desno ter navzgor. Slabost te zbirke je, da ni v celoti javno dostopna v raziskovalne namene. V tem času se raziskovalci trudijo izboljšati trenutno stanje tovrstne tehnologije z metodami globokega učenja [45, 46].

2.3.2. Razpoznavanje na podlagi žil beločnice

Večina raziskovalcev razpoznavalne algoritme na podlagi beločnice razvija na sledečih podatkovnih zbirkah: UBIRIS.v1 [47], UBIRIS.v2 [48], UTIRIS [49], MICHE-I [50] ter MASD [20]. Zbirke so tipično primerne za prepoznavanje na podlagi beločnice, če so slike zajete v vidnem spektru. Pri mnogih zbirkah [6], ki so namenjene prepoznavanju na podlagi šarenice, po protokolu zajemajo slike v spektru, ki je blizu infrardečega (angl. near infrared, NIR). Zato niso primerne za prepoznavanje na podlagi beločnice, saj v tem spektru večino informacije o beločnici izgubimo. Cevovod za razpoznavanje na podlagi beločnice ponavadi sestavlja naslednje visokonivojske faze [38, 40, 51]:

- Segmentacija beločnice (angl. sclera segmentation).
- Izboljšanje informacije vaskularnih struktur (angl. vessel enhancement).
- Izluščevanje značilnk (angl. feature extraction).
- Izračun ujemanja (angl. matching).

Informacijo o vaskularnih strukturah moramo izboljšati zato, da izboljšamo kvaliteto vzorca žil. Če tega koraka ne naredimo, vaskularne strukture zaradi pomanjkanja kontrasta pogosto ne izstopajo dovolj in jih nadaljnji algoritmi zato težje zaznajo. V literaturi sta najbolj pogosto uporabljena algoritma za izboljšanje vaskularnih struktur kontrastno-omejena adaptivna izenačitev histograma (angl. contrast-limited adaptive histogram equalization, CLAHE) in obogatitev z Gaborjevimi filtri [22, 52]. Za fazo izluščevanja značilnk je bilo predlaganih mnogo različnih metod, na primer: značilke na osnovi valjev (angl. wavelet-based features) [41], GLCM [53], gosti SIFT [54], gosti LBP [55] in OLBP [51]. Za mnogo algoritmov so v sklopu SSERBC [20] tekmovanja narejene primerjalne analize, vendar je na tem mestu potrebno poudariti, da je v nekaterih izmed teh algoritmov (poleg informacije iz beločnice) uporabljena tudi informacija iz drugih predelov očesa.

2.4. Razpoznavanje na podlagi periokularnega dela

Periokularna oziroma obočesna regija vključuje površino obraza, ki je v neposredni bližini očesa (veke, obrvi, zgornji del lic). V delu [56] je bila predstavljena ideja, da to

regijo lahko uporabljamo za razpoznavalne namene kot dopolnilno biometrično modalnost šarenici (povzeto po [57, 58]). Lahko jo uporabljamo tudi kot samostojno biometrično modalnost [16]. Prednost periokularne regije je, da jo pogosto lahko zajamemo tudi, kadar je protokol za zajem manj omejen (angl. unconstrained). Kakor ugotavljajo avtorji [59], si z njo lahko pomagamo v scenarijih, ko kvaliteten zajem šarenice ni možen (npr. ko subjekt ni kooperativen, oko je zaprto itd.).

Za raziskave razpoznavanja na podlagi periokularnega dela lahko uporabljamo podatkovne zbirke, ki se osredotočajo na razpoznavanje obrazov (npr. [60, 61]) ali šarenice (npr. [47, 62, 63]). Prav tako obstajajo podatkovne zbirke, namenjene izključno raziskavam periokularnega dela [57, 64].

Literatura v zvezi s periokularnim razpoznavanjem je v grobem razdeljena na dva dela: *i*) opis tehnik, ki se ukvarjajo z zaznavo in segmentacijo periokularne regije in *ii*) opis tehnik za izluščevanje diskriminativnih značilk iz periokularnega dela. Natančnejši pregled tehnik iz prve skupine je opisan v delu [58], nas pa bodo večinoma zanimala tehnike iz druge skupine. Poznamo dve družini algoritmov za izluščevanje značilk, glede na to, kakšen obseg slike v procesu izluščevanja upoštevajo, in sicer *i*) globalne algoritme (barva, oblika, tekstura) ter *ii*) lokalne algoritme. V delu [56] za izluščevanje značilk naredijo eksperimente razpoznavanja z dvema globalnima algoritmoma in enim lokalnim algoritmom. Uporabljena globalna algoritma sta algoritem lokalnih binarnih vzorcev (angl. local binary pattern, LBP) in algoritem histograma usmerjenih gradientov (angl. histogram of oriented gradients, HOG), uporabljen lokalni algoritem pa je algoritem transformacije na skaliranje invariantnih značilk (angl. scale-invariant feature transform, SIFT). Ker definicija periokularnega dela lahko zajema oziroma ne zajema obrvi, avtorji v nadaljevanju prvotnega dela [65] naredijo tudi študijo o tem, kako vsebovanost obrvi vpliva na uspešnost razpoznavanja. Periokularni del lahko (za razliko od same šarenice in vaskularnih struktur) razkrije tudi drugačne vrste informacije, kot je na primer spol in etnična pripadnost [66]. V delu [67] lahko najdemo nedavno narejen pregled uspešnosti različnih algoritmov za razpoznavanje na podlagi periokularnega dela.

2.5. Razpoznavanje na podlagi šarenice

Razpoznavanje na podlagi šarenice je sestavljeno iz cevovoda, ki tipično zajema naslednje faze [68]:

- Lokalizacija očesa na sliki.
- Segmentacija šarenice in zenice (zaznavanje mej).
- Normalizacija.

- Izluščevanje značilik (angl. feature extraction).
- Izračun ujemanja (angl. matching).

Poznamo več podatkovnih zbirk [69], ki so namenjene raziskavam razpoznavanja na podlagi šarenice, kot so na primer CASIA [62], UBIRIS [47] ter UPOL [63]. Algoritem IrisCode [5], ki ga je razvil Daugman leta 1993, v mnogih raziskavah [70, 71, 72, 73] služi kot izhodišče za razpoznavanje na podlagi šarenice. Daugman je za segmentacijo šarenice iz slike celotnega očesa najprej predlagal zaznavo robov z integro-diferencialnim operatorjem. Algoritem zazna notranjo mejo šarenice, ki meji na zenico, ter zunanjo mejo šarenice, ki meji na beločnico. Slabost integro-diferencialnega operatorja je, da predpostavlja, da sta obe meji šarenice krožne oblike, kar je precejšnja omejitev, še posebej kadar imamo opravka z manj omejenimi protokoli zajema slik. Za naslovitev tega problema je leta 2007 Daugman predlagal odpornejšo segmentacijsko metodo z aktivnimi konturami [7]. Faza normalizacije temelji na transformaciji, ki šarenico transformira iz kartezičnih koordinat v psevdo polarne koordinate, s čimer dobimo kvadratno predlogo. Na podlagi te pravokotne predloge (angl. template) lahko računamo ujemanje med dvema šarenicama.

V zadnjih letih se raziskovalci lotevajo prepoznavanja na podlagi šarenice z metodami globokega učenja [35, 74], ki so se (kakor bomo videli v nadaljevanju tega dela) na področju računalniškega vida velikokrat izkazale kot najbolj uspešne. Avtorji dela [33] pokažejo, da je doseganje dobrih razpoznavalnih rezultatov z globokim učenjem možno. Zato, da z globokim učenjem lahko dosežemo dobre rezultate, ponavadi potrebujemo veliko količino podatkov. V omenjenem delu je izpostavljena tudi težava, da je za prepoznavanje na podlagi šarenice v javni domeni trenutno še vedno premalo podatkov, da bi modele lahko učili brez prej naučenih modelov na drugih domenah. Zato je predlagano, da si pomagamo z metodami prenešenega učenja (angl. transfer learning).

2.6. Biometrična fuzija očesnih modalnosti

Biometrična fuzija pomeni, da v sistemu za razpoznavanje združimo informacijo iz različnih biometričnih modalnosti. Fuzija večih modalnosti ima potencial izboljšati natančnost razpoznavanja, ki bi jo sicer dobili v eno-modalnem sistemu. Glede na tip združevanih modalnosti poznamo dve vrsti okularnih biometričnih fuzij [12], in sicer: *i*) intra-okularno fuzijo ter *ii*) fuzijo okularnih modalnosti skupaj z drugimi (ne-očesnimi) modalnostmi. Pri intra-okularni fuziji združujemo le informacijo, pridobljeno iz različnih okularnih modalnosti. V tem delu nas bo zanimal izključno intra-okularni tip fuzije, saj se bomo ukvarjali z združevanjem informacije iz šarenice, vaskularnih struktur in periokularnega dela.

Glede na to, na katerem nivoju združujemo informacijo večih modalnosti, ločimo različne vrste fuzije. Informacijo lahko združujemo takoj zatem, ko senzor zajame modalnost (angl. sensor level). Pred izračunom ujemanja lahko združujemo tudi značilke posameznih modalnosti (angl. feature level). Po izračunu ujemanja lahko informacijo združujemo na treh nivojih, in sicer: *i*) na nivoju ocene (angl. score level), *ii*) na nivoju ranga (angl. rank level) ali *iii*) na nivoju odločitve (angl. decision level). V sorodnih delih, ki so povzeta v [12], se za fuzijo očesnih modalnosti večinoma uporablja združevanje na nivoju ocene. V delu [75] so izboljšali razpoznavanje na podlagi šarenice tako, da so barvno sliko šarenice obravnavali ločeno po posameznih barvnih kanalih (R,G in B) ter za pridobitev končne napovedi združili razpoznavalne rezultate posameznih kanalov. Fuzijo so naredili na nivoju ocene. V delu [76] primerjajo uspešnost štirih različnih pristopov za razpoznavanje na podlagi šarenice, primerjani pristopi pa so: diskretna kosinusna transformacija (angl. discrete cosine transform, DCT), hitra Fourierjeva transformacija (angl. fast Fourier transform, FFT), Log-Gabor in Haar valjčno transformacijo (angl. Haar wavelet transform, HWT). Avtorji tudi primerjajo uspešnost različnih fuzij teh pristopov. Gre ponovno za fuzijo na nivoju ocene, pri čemer rezultati pokažejo, da najbolje deluje utežena vsota Log-Gabor in Haar wavelet algoritma. V delu [39] naredijo fuzijo šarenice in vaskularnih struktur (na nivoju ocene), pri čemer fuzija obeh modalnosti izboljša rezultat. V delu [77] pokažejo, da lahko fuzija šarenice in periokularnega dela izboljša natančnost razpoznavanja.

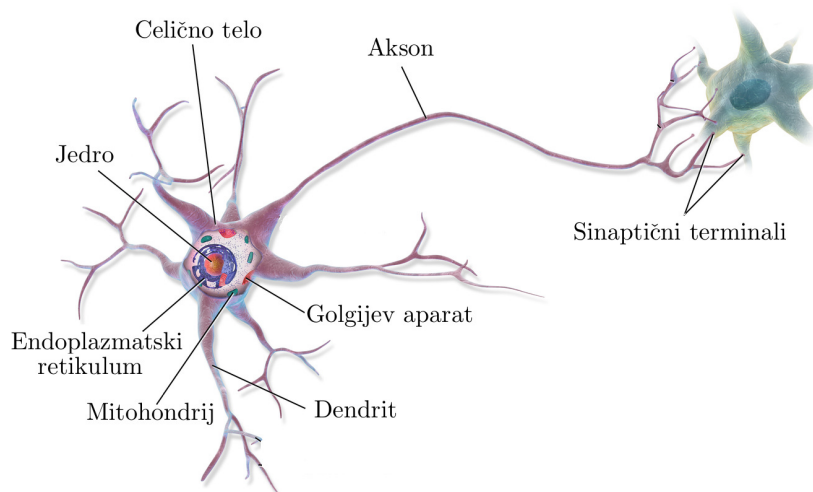
Poglavje 3

Nevronske mreže

3.1. Osnovni gradniki nevronske mreže

3.1.1. Biološke nevronske mreže

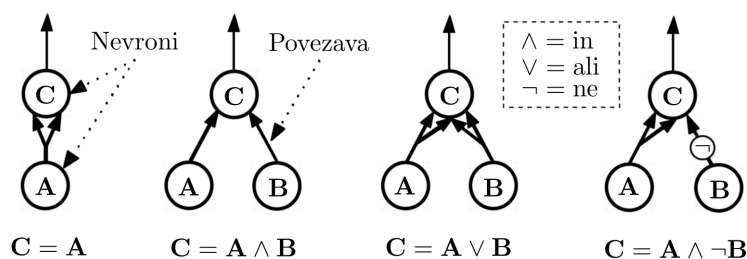
Ideja za matematično definicijo umetne nevronske mreže (angl. artificial neural network, ANN) izhaja iz bioloških nevronske mreže (angl. biological neural network, BNN), ki se med drugim nahajajo v človeških možganih in hrbtenjači. Osnovni gradnik biološke nevronske mreže so živčne celice oziroma nevroni [78]. Nevroni so specializirane celice za sprejem, procesiranje ter prenos električnih in kemičnih signalov. Osnovna zgradba nevrona je predstavljena na sliki 3.1. Na kratko bomo predstavili dele biološkega nevrona, ki bodo kasneje preslikani v matematično obliko. Deli biološkega nevrona so: dendriti (angl. dendrites), akson (angl. axon) ter celično telo, ki mu rečemo tudi soma (angl. cell body / soma). Iz telesa nevrona se izraščata več dendritov, na katere se lahko veže več aksonov drugih nevronov. Isti nevron se s svojim aksonom lahko naprej veže na druge nevrone v biološkem sistemu – na tak način nevroni tvorijo mreže. Posamezen nevron večino signalov, ki mu jih pošljejo drugi nevroni, prejme preko dendritov. Te signali so nato interpretirani v celičnem telesu. Če je signal pridobljen iz dendritov dovolj močan, nevron pošlje signal, imenovan aktivacijski potencial (angl. action potential), vzdolž svojega aksona ter sinaptičnih terminalov drugim nevronom. Seveda je to le poenostavljena predstavitev delovanja, ki pa bo zadostovala za nadaljnje razumevanje.



Slika 3.1: Zgradba biološkega nevrona [79].

3.1.2. Umeten nevron

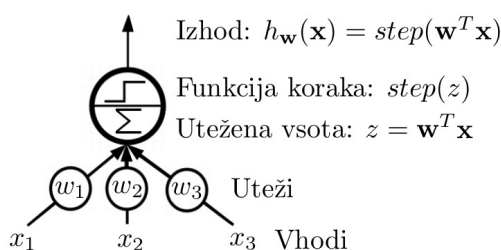
Na podlagi biološkega modela, ki smo ga opisali, sta Warren McCulloch in Walter Pitts [80] leta 1943 predlagala umeten nevron (angl. artificial neuron). Povezanost nevronov sta predstavila kot graf, kjer vozlišča predstavljajo somo z dendriti (na katere se lahko vežejo drugi nevroni), aksoni pa so preslikani v povezave med vozliči. Umeten nevron lahko na vhod prejme le binarne vhode, izhodi pa so lahko le binarne vrednosti. S predlaganim umetnim nevronom lahko na preprost način realiziramo identiteto in nekaj osnovnih logičnih funkcij, kot je prikazano na sliki 3.2. Nevron je aktiviran, če sta vsaj 2 vhoda aktivirana.



Slika 3.2: Realizacija osnovnih logičnih funkcij z umetnim nevronom [81].

3.1.3. Linearna pragovna enota

Linearna pragovna enota (angl. linear threshold unit, LTU) se od umetnega nevrona razlikuje po tem, da na vhod namesto binarnih vrednosti damo poljubna števila. Na sliki 3.3 so vhodi v LTU predstavljeni s spremenljivkami x_1, x_2, x_3 (v splošnem jih je lahko poljubno mnogo). Vsak vhod ima tudi svojo korespondenčno utež w_1, w_2, w_3 .



Slika 3.3: Linearna pragovna enota, LTU [81].

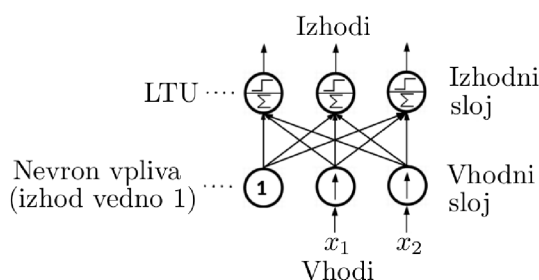
LTU izračuna uteženo vsoto vhodov, ki jo lahko zapišemo v obliki skalarnega produkta kot $z = \mathbf{w}^T \cdot \mathbf{x}$. Izračuna se še tako imenovana koračna funkcija – dve tipični koračni funkciji sta predstavljeni z enačbama (3.1) in (3.2). Rezultat koračne funkcije gre na izhod LTU.

$$heaviside(z) = \begin{cases} 0, & \text{če } z < 0 \\ 1, & \text{če } z \geq 0 \end{cases} \quad (3.1)$$

$$sgn(z) = \begin{cases} -1, & \text{če } z < 0 \\ 0, & \text{če } z = 0 \\ 1, & \text{če } z > 0 \end{cases} \quad (3.2)$$

3.1.4. Enoslojni perceptron

Leta 1957 je Frank Rosenblatt [78, 82] razvil perceptron (angl. single layer perceptron), ki temelji na LTU. Na tem mestu je potrebno omeniti, da je v literaturi definicija perceptrona osnovana bodisi le na enem LTU (na ta primer se navezuje naša nadaljnja razlaga) ali pa na sloju le teh (prikazano na sliki 3.4).



Slika 3.4: Enoslojni perceptron [81].

Pri perceptronu se vsakemu LTU na vhod doda še nevron vpliva (angl. bias neuron). Pri biološkem nevronu smo rekli, da se sproži signal (aktivacijski potencial) le, če je skupen signal, ki ga nevron pridobi preko dendritov, dovolj močan (preseže določen prag). Ta prag bomo pri perceptronu označili s spremenljivko t (angl. threshold), primerjali pa ga bomo z vsoto vseh uteženih vhodov, kakor prikazuje enačba perceptrona (3.3).

$$\text{izhod} = \begin{cases} 1, & \text{če } \sum_j \mathbf{w}_j \mathbf{x}_j \geq t \\ 0, & \text{drugače} \end{cases} \quad (3.3)$$

To enačbo lahko na drugačen način zapišemo tako, da vsoto produktov uteži in vhodov zapišemo kot skalarni produkt, prag t pa prestavimo na drugo stran enačbe. Prag t bomo v istem koraku še preimenovali v *vpliv*, kakor je prikazano v enačbi (3.4). S spremenljivko *vpliv* uravnavamo, kako lahko je doseči, da je na izhodu perceptrona visoko stanje (večji kot je *vpliv*, lažje je ob danih vhodih in utežeh doseči 1 na izhodu).

$$\text{izhod} = \begin{cases} 1, & \text{če } \mathbf{w}^T \cdot \mathbf{x} + \text{vpliv} \geq 0 \\ 0, & \text{drugače} \end{cases} \quad (3.4)$$

Perceptron lahko učimo reševati nek problem s posodabljanjem uteži glede na napako perceptrona pri napovedovanju, kakor je prikazano z enačbo (3.5). η je stopnja učenja (angl. learning rate) in člen v oklepaju razlika med pravo ciljno vrednostjo in vrednostjo, ki jo napove perceptron.

$$\mathbf{w}_{i,j}^{(\text{naslednji korak})} = \mathbf{w}_{i,j} + \eta(y_j - \hat{y}_j)\mathbf{x}_i \quad (3.5)$$

Enojen sloj perceptronov ima dve omejitvi in sicer:

1. Rešuje le linearno ločljive probleme (ne more se naučiti bolj kompleksnih funkcij).
2. Na izhode poda le binarne vrednosti in ne verjetnosti (kakor na primer logistična regresija). To ni dobro zato, ker ob majhnih spremembah uteži lahko pride do

popolne negacije izhodne vrednosti. Za proces učenja bi imeli rajši bolj predvidljiv izhod – ob majhnih spremembah uteži, sorazmerno majhno spremembo izhodne vrednosti.

3.1.5. Sigmoidni nevron

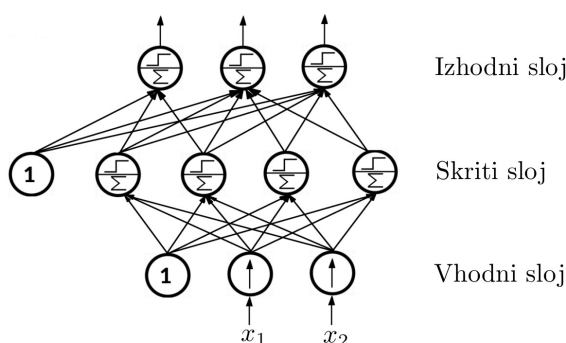
Sigmoidni nevron se od klasičnega perceptrona razlikuje po tem, da namesto tipične koračne funkcije uporabimo sigmoidno funkcijo, ki je predstavljena z enačbo (3.6). Spremenljivka z (enako kot pri perceptronu) opisuje skalarni produkt vhodov in njihovih uteži. Upoštevan je tudi *vpliv*, torej $z = \mathbf{w}^T \cdot \mathbf{x} + \text{vpliv}$.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.6)$$

Izhod sigmoidnega nevrona je vrednost med 0 in 1 (torej ne gre več le za binarno vrednost kot pri navadnem perceptronu). Sigmoidni nevron uvedemo, ker nam omogoča, da v sistem (kakor bomo videli kasneje) uvedemo učinkovite algoritme za vzvratno razširjanje napake (angl. backpropagation). To prej ni bilo mogoče, saj izhod perceptrona ni bil dovolj predvidljiv glede na majhne spremembe uteži, kot smo opisali v poglavju 3.1.4.

3.1.6. Večslojni perceptron

Večslojni perceptron (angl. multi-layer perceptron, MLP) označuje več zaporednih slojev nevronov. Čeprav k temu namiguje ime, ni nujno, da gre za klasične perceptrone s koračno funkcijo – nevroni imajo lahko tudi drugo aktivacijsko funkcijo. Primer večslojnega perceptrona je prikazan na sliki 3.5.



Slika 3.5: Večslojni perceptron [81].

Večslojni perceptron mora imeti po definiciji vsaj 3 nivoje vozlišč: vhodni sloj, skriti sloj (angl. hidden layer) ter izhodni sloj. Če ima dva ali več skritih slojev, gre za globoko nevronska mrežo (angl. deep neural network, DNN). Vozlišča enega sloja morajo biti povezana z vsemi vozlišči naslednjega sloja. Kot težavo enoslojnega perceptrona smo omenili, da lahko rešuje le linearno ločljive probleme, večslojni perceptron pa zaradi večih nivojev naslovi to težavo in se je zmožen naučiti tudi bolj kompleksnih funkcij.

3.1.7. Aktivacijske funkcije

V poglavju 3.1.5 smo videli, da pri perceptronu koračno funkcijo lahko zamenjamo s sigmoidno funkcijo, ki spada v družino aktivacijskih funkcij. V splošnem jo lahko zamenjamo s katero koli funkcijo iz te družine. Aktivacijska funkcija je tisti del nevrona, ki odloča, kolikšna bo izhodna vrednost tega nevrona (glede na skalarni produkt uteži in vhodov). Če potegnemo vzporednico z biološkim sistemom (opisano v poglavju 3.1.1), bi lahko rekli, da aktivacijska funkcija odloča o količini aktivacijskega potenciala (od tod tudi ime te družine funkcij). V literaturi večkrat zasledimo tudi izraz prehodna funkcija (angl. transfer function), ki se pogosto uporablja kot sinonim za aktivacijsko funkcijo.

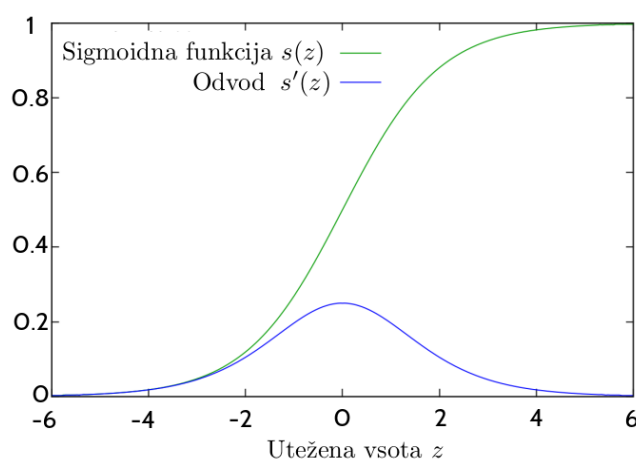
Ena izmed prvih idej je, da bi koračno funkcijo lahko zamenjali s preprosto linearno funkcijo. Poglejmo si razloge, zakaj izbira ni primerna:

1. Če si predstavljamo večslojno mrežo, pri kateri ima vsak nevron linearen izhod glede na linearno kombinacijo vhodov (t.j. skalarni produkt vhodov in uteži), potem se celotna mreža (kljub večjemu številu nivojev) lahko nauči le linearnih funkcij.
2. Kakor bomo videli v poglavju 3.1.9, je za aktivacijsko funkcijo pomembno tudi to, kakšen je njen odvod – na podlagi odvoda izboljšujemo uteži. Ker vemo, da je odvod linearne funkcije konstanta (v odvodu niso zastopane uteži), mreže ne bomo mogli učinkovito učiti.
3. Izhod linearne funkcije ni omejen, gre za interval $(-\infty, +\infty)$.

Sigmoidna funkcija odpravlja vse omenjene slabosti linearne funkcije (je nelinearna, ima ustrezen zvezen odvod ter omejen interval $[0, 1]$). Kljub temu pa ima svoje slabosti, in sicer:

1. Če je vhodna utežena vsota $z = 0$, na izhodu dobimo pozitivno vrednost (aktivacija z vrednostjo 0,5), pri čemer bi bilo bolj intuitivno, da aktivacije ne bi bilo.
2. Če je vhodna utežena vsota izjemno majhna (ali velika), je funkcijska vrednost zelo blizu 0 (ali 1). Ob taki situaciji pravimo, da je prišlo do zasičenja sigmoidne funkcije (angl. saturation). Ko postopek vzvratnega popravljanja napake izračuna

odvode sigmoide, so odvodi zelo majhni (vrednost je blizu 0). Funkcijo sigmoide z njenim odvodom prikazuje slika 3.6. Če so ob tem pojavu vhodne vrednosti izjemno majhne, to imenujemo problem izginjajočih gradientov (angl. vanishing gradients problem). Če so vhodne vrednosti izjemno visoke, temu pravimo, da gradienti eksplodirajo (angl. exploding gradients problem).



Slika 3.6: Graf sigmoide in njenega odvoda.

Za naslovitev prve opisane težave sigmoidne funkcije je predlagana funkcija hiperbolični tangens $\tanh(z)$, ki jo prikazuje enačba (3.7), pri čemer z ponovno označuje uteženo vsoto vhodov, e pa Eulerjevo število. Kot vidimo, jo lahko interpretiramo kot skalirano in (glede na y os) centrirano sigmoidno funkcijo σ . Izhodne vrednosti zavzamejo interval $[-1, 1]$. Očitno pa je, da ima \tanh še vedno težave z zasičenjem.

$$\tanh(z) = \frac{2}{1 + e^{-2z}} - 1 = 2 * \sigma(2z) - 1 \quad (3.7)$$

Popravljen linearna enota (angl. rectified linear unit, ReLU) je preprosta aktivacijska funkcija, ki odpravlja težave z zasičenjem. Opisana je z enačbo (3.8). Njen izhodni interval je $[0, +\infty)$. V primerjavi s sigmoido in \tanh je tudi hitreje izračunljiva in aktivacije so bolj redke (angl. sparse), kar je posledica tega, da se pri vseh negativnih vhodnih vrednostih nevron ne aktivira [83].

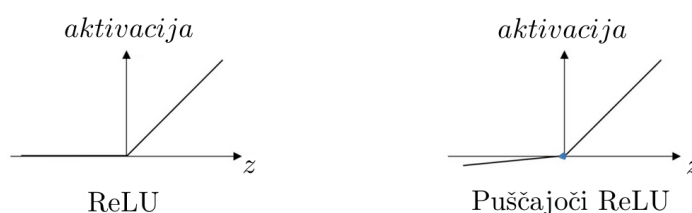
$$\text{relu}(z) = \max(0, z) \quad (3.8)$$

Problem ReLU je tako imenovano umiranje nevronov (angl. dying ReLUs), kar pomeni, da nevroni na izhod začnejo dajati izključno vrednost 0. To se kljub popraviljanju uteži v fazi vzratnega popraviljanja napake ne spremeni, saj je gradient ReLU pri negativni

vrednosti enak 0. Zato je bila predlagana funkcija puščajoči ReLU (angl. leaky ReLU), ki omogoča, da nevroni spet oživijo. Puščajoči ReLU je predstavljen z enačbo (3.9). Pri tem je α majhna vrednost, npr. 0,01.

$$\text{puščajoči_relu}(z) = \begin{cases} \alpha * z & \text{če } z < 0 \\ z, & \text{drugače} \end{cases} \quad (3.9)$$

ReLU in puščajoči ReLU sta prikazani na sliki 3.7.



Slika 3.7: Primerjava aktivacijskih funkcij ReLU in puščajoči ReLU.

3.1.8. Napovedovanje in merjenje napake

Do zdaj smo v delu predstavili strukturo aciklične nevrnske mreže (angl. feed-forward neural network, FNN). Če mreža tvori cikle, jo imenujemo rekurzivna nevrnska mreža (angl. recursive neural network), vendar nas tak primer mreže v tem delu ne bo zanimal. RNN se namreč tipično uporablja pri obdelavi podatkov, kjer so pomembne relacije med elementi zaporedja (npr. pri obdelavi videa gre za relacije med zaporednimi slikami, pri glasbi pa med zaporednimi zvoki). S FNN ne moremo modelirati podobnih časovnih odvisnosti.

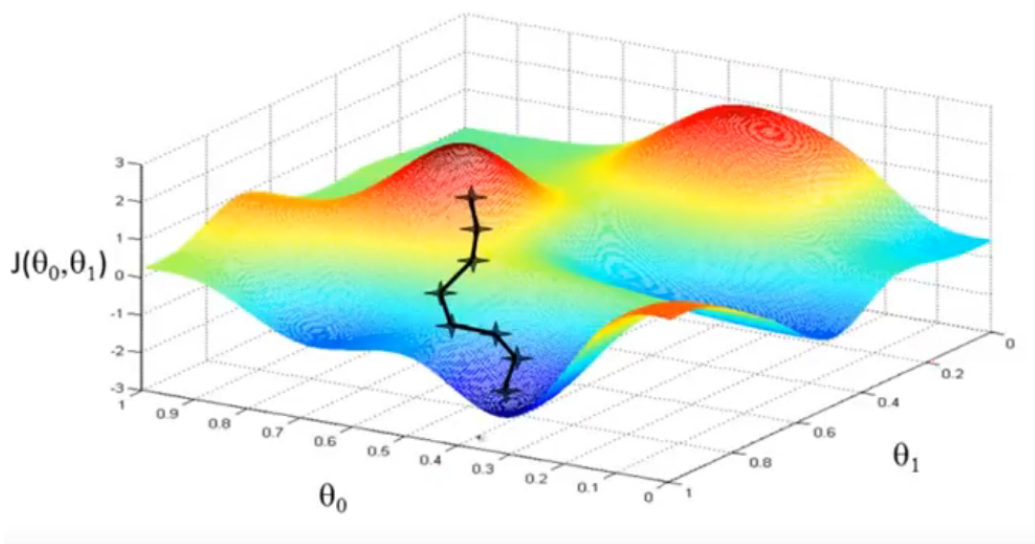
Glavni cilj učenja nevrnske mreže je doseči čimbolj natančno preslikavo $\mathbf{X} \rightarrow \mathbf{Y}$, pri čemer so elementi množice \mathbf{X} vhodni podatki in elementi množice \mathbf{Y} ciljne spremenljivke (angl. target variables / groundtruth). To preslikavo naredimo s CNN modelom, za katerega moramo najti ustrezne uteži \mathbf{W} . Natančneje pri učenju želimo za model najti množico uteži \mathbf{W} , ki minimizirajo cenilno funkcijo (angl. cost function) $J(\mathbf{W})$. Preslikava bo torej natančna, če bomo našli ustrezne učljive parametre (angl. learnable parameters), ki so v našem primeru uteži. Nevroni vpliva se v literaturi prav tako uvrščajo med učljive parametre, čeprav s stališča implementacije ta vozlišča na izhod dajejo le eno konstantno vrednost. Ker ob učenju nevrnske mreže pri nevronih vpliva spreminjamo oz. učimo le utež na povezavi, ki vodi iz nevrna vpliva, bodo v nadaljevanju tega dela ti nevroni vsebovani v pojmu uteži.

Mrežo učimo z učno množico (angl. *trainset*), ki vsebuje pare elementov (\vec{x}_i, y_i) , kjer $\vec{x}_i \in \mathbf{X}$ in $y_i \in \mathbf{Y}$. V naši notaciji je \vec{x}_i vektor značilik i -tega učnega primera, y_i pa razred, ki mu i -ti primer pripada. Pri problemih klasifikacije elementi množice \mathbf{Y} pripadajo končni množici diskretnih spremenljivk, ki jim rečemo razredi. Pri problemih regresije je ciljna spremenljivka zvezna.

Predstavljamo si, da mreži podamo na vhod par (\vec{x}_i, y_i) . Prenajanje naprej (angl. *forward-pass / inference*) označuje proces napovedovanja nevronske mreže glede na vhod \vec{x}_i in trenutno konfiguracijo ucljivih parametrov. Na izhodu nevronske mreže dobimo napoved \hat{y} , ki na začetku učenja nevronske mreže tipično odstopa od ciljne spremenljivke y_i . Napako napovedi merimo s tako imenovanimi funkcijami izgube (angl. *loss functions*).

3.1.9. Gradientni spust

Gradientni spust je optimizacijski algoritem, ki iterativno minimizira cenovno funkcijo $J(\mathbf{W})$ pri podanih utežeh \mathbf{W} [81, 84]. Predpostavimo, da imamo samo dve uteži θ_1 in θ_2 , in želimo optimizirati funkcijo $J(\theta_1, \theta_2)$. Funkcijo $J(\theta_1, \theta_2)$ s prostorom vseh rešitev lahko grafično predstavimo na podoben način, kakor je prikazano na sliki 3.8. Predpostavimo tudi, da se naš začetni krajevni vektor nahaja na sredini rdečega hriba.



Slika 3.8: 2D graf cenilne funkcije J pri poljubnih parametrih θ_1 in θ_2 [84].

Vemo, da gradient krajevnega vektorja, ki ga sestavljajo parcialni odvodi vseh spremenljivk, kaže v smeri največjega naraščanja funkcije, glede na krajevni vektor. V želji, da

bi konvergirali proti globalnemu minimumu funkcije, se moramo torej premikati v nasprotni smeri gradienta. V primeru, prikazanem na sliki 3.8, se v sedmih korakih premaknemo v globalni minimum funkcije, ki je označen s temno modro barvo. S členom α , ki ga imenujemo stopnja učenja (angl. learning rate), definiramo, kako velike korake bomo delali v nasprotni smeri gradienta. Iterativni algoritem gradientnega spusta je predstavljen s psevdokodo:

```
ponavljaj do konvergence {
    // izračunaj desno stran enačbe in simultano posodobi
    // vse uteži theta
     $\theta_j := \theta_j - \alpha \frac{\delta}{\delta \theta_j} J(\theta_0, \theta_1)$ 
}
```

Glede na način podajanja podatkov poznamo tri vrste gradientnega spusta, s katerimi posodabljammo uteži [81]:

1. Paketni gradientni spust (angl. batch gradient descent) za eno samo posodobitev izračuna gradient celotne podatkovne zbirke. Slabost tega je, da je izračun počasen, prav tako pa smo omejeni s spominom računalnika (če je podatkovna zbirka prevelika, je izračun nemogoč). Ker ponavadi dobljenega modela ne moremo izboljševati v trenutku, ko v svojo bazo dobimo nov primer, tovrstno učenje imenujemo učenje brez povezave (angl. offline learning).
2. Stohastični gradientni spust (angl. stochastic gradient descent, SGD) za posodobitev uporabi en sam učni primer. Posodobitve so zato hitrejše, vendar je konvergenca modela k rešitvi ponavadi slabša v primerjavi z drugimi vrstami gradientnega spusta, saj so optimizacijski koraki narejeni na občutno manj podatkih. Ker posodobitve lahko hitro računamo, ko je naš model v produkcijskem okolju, tako vrsto učenja imenujemo učenje s povezavo (angl. online learning).
3. Gradientni spust z majhnimi paketi (angl. mini-batch gradient descent) deluje tako, da uteži posodabljammo glede na manjše skupine učnih podatkov. Konvergenca je ponavadi hitrejša kot pri SGD, čas izračuna pa le malo daljši, zato je to priljubljena vrsta gradientnega spusta, ki ga bomo v tem delu uporabljali tudi mi.

Za popravljanje napake je potrebno izračunati gradiente funkcije izgube in našega modela. Gradiente lahko učinkovito izračunamo z algoritmom vzvratnega razširjanja napake (angl. backpropagation algorithm) [85].

3.2. Konvolucijske nevronske mreže

3.2.1. Uvod v konvolucijske nevronske mreže

Konvolucijske nevronske mreže (angl. convolutional neural networks, CNNs / ConvNets) so vrsta acikličnih nevronskih mrež, ki posnemajo delovanje vidnega korteksa [81, 86]. Hubel in Wiesel sta s svojimi eksperimenti na mačkah ugotovila [87], da so celice v vidnem korteksu specializirane za zaznavo določenih oblik. Pokazala sta na primer, da se nekatere skupine celic bolj vzdražijo, če so v vidnem polju horizontalne črte, medtem ko se druge bolj odzivajo na vertikalne črte. Zaradi razvoja konvolucijskih mrež so bili na področju računalniškega vida narejeni veliki napredki pri reševanju nalog iz klasifikacije [29], mreže za klasifikacijo pa se v predelani obliki, kakor bomo videli v nadaljevanju, da uspešno uporabiti tudi za segmentacijo.

3.2.2. Arhitektura konvolucijskih nevronskih mrež

Prva arhitekturna razlika, po kateri se CNN od klasične nevronske mreže razlikuje je, da na vhod ne dajemo več le enodimenzionalnih podatkov – učne primere smo v prejšnjem poglavju namreč opisovali z 1D vektorjem \vec{x} . Učni podatki pri CNN (v našem delu bo šlo za slike) imajo najmanj dve dimenziji – črno-belo sliko definiramo kot 2D polje slikovnih elementov, RGB slike pa imajo še dodatno dimenzijo globine. CNN tipično sestavljajo naslednji sloji [88, 89]:

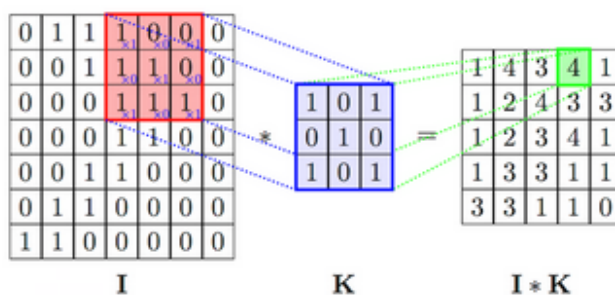
1. Konvolucijski sloj (angl. convolutional layer).
2. ReLU sloj (angl. ReLU layer).
3. Združevalni sloj (angl. pooling layer / downsampling layer).
4. Polno-povezani sloj (angl. fully-connected layer).

Konvolucijski sloj

Konvolucijski sloj se v CNN arhitekturi ponavadi pojavi kot prvi sloj. V mreži se tipično pojavi več takih slojev. Za izračun konvolucije nad 2D sliko I (dimenzij $w \times h$) potrebujemo majhno 2D matriko K (dimenzij $w_k \times h_k$, pri čemer $w_k \ll w$ in $h_k \ll h$), ki ji rečemo jedro konvolucije (angl. convolution kernel). V literaturi se kot sinonim za konvolucijsko jedro uporabljata tudi izraza filter in detektor značilk (angl. feature detector). Jedro K

pomikamo po sliki I , kakor je prikazano na sliki 3.9. Ob vsakem premiku izračunamo vsoto sosežnih elementov delca slike (dimenzije $w_k \times h_k$) in jedra K , kakor prikazuje enačba 3.10. Elementom matrike I , ki so aktualni glede na trenutno pozicijo jedra (na sliki 3.10 označeni z rdečo barvo), imenujemo tudi sprejemno polje (angl. receptive field). Vsote zapisujemo v matriko, ki jo poimenujemo mapa aktivacij oziroma mapa značilnk (angl. activation map / feature map). Le-ta predstavlja končni rezultat konvolucije, ki jo bomo označili $I * K$.

$$(I * K)_{xy} = \sum_{i=1}^{h_k} \sum_{j=1}^{w_k} K_{ij} * I_{x+i-1, y+j-1} \quad (3.10)$$



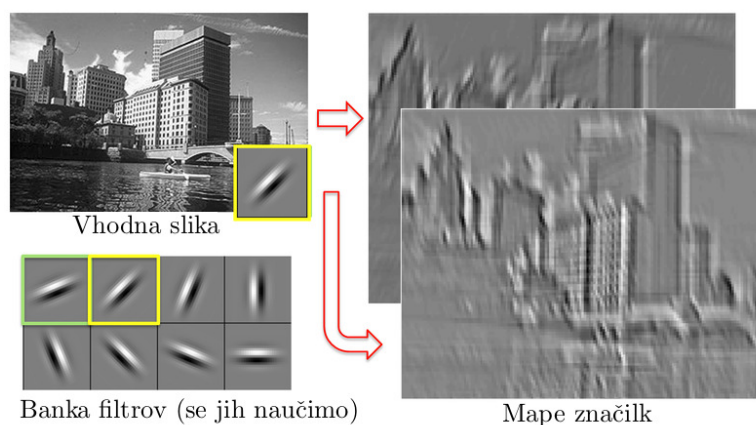
Slika 3.9: Premikanje konvolucijskega jedra pri računanju konvolucije [89].

Elementi konvolucijskih jeder so torej ucljivi parametri (uteži), ki jih med učenjem nevronske mreže optimiziramo. V splošnem velja, da mora biti globina jedra enaka globini vhoda – če bi bil vhod slika s tremi kanali, mora biti tudi jedro (ki ga premikamo po tej sliki) veliko $w_k \times h_k \times 3$.

Elementi matrike I v CNN so ponavadi slikovni elementi vhodne slike (če se nahajamo na začetku mreže) ali izhodne aktivacije nevronov (če gledamo konvolucijski sloj, ki je v sredi/pri koncu mreže). Element (i, j) v matriki I torej predstavlja izhod nevrona (i, j) . Element (i, j) v matriki $I * K$ pa pomeni uteženo vsoto (ki smo jo prej zapisali kot z), ki pride na vhod nevrona (i, j) . V splošnem ne moremo reči, da so elementi $I * K$ že izhodne vrednosti nevronov, saj manjka še aktivacijska funkcija.

Videli smo, da se elementi I (oz. izhodne vrednosti nevronov), ki ležijo na sprejemnem polju, preko ucljivih uteži preslikajo v le en element na sliki aktivacij $I * K$. En element v matriki $I * K$ predstavlja mnogo elementov iz matrike I . En nevron torej zastopa svojo lokalno okolico, čemur rečemo tudi lokalna povezanost (angl. local connectivity).

Do zdaj smo opisovali sistem, ki ima pri enem sprejemnem polju le en filter, moramo pa se zavedati, da v običajni CNN arhitekturi za vsako sprejemno polje učimo množico različnih filtrov, kakor je prikazano na sliki 3.10.



Slika 3.10: Za vsako sprejemno polje učimo množico uteži pri različnih filtrih [90].

Ko govorimo o konvolucijskih filtrih, je pomembno razumeti tudi, kaj filtri semantično opisujejo, glede na to, kako globoko v mreži se nahajajo. V delih, ki se ukvarjajo z vizualizacijo filtrov (na primer [91]), je bilo pokazano, da se v prvih nivojih mreža nauči enostavnih, nizkonivojskih značilk, kot so na primer orientirani robovi. Bolj, kot je opazovani nivo mreže globoko, bolj postanejo značilke abstraktne, kar interpretiramo tako, da kodirajo višjenivojske koncepte.

ReLU sloj

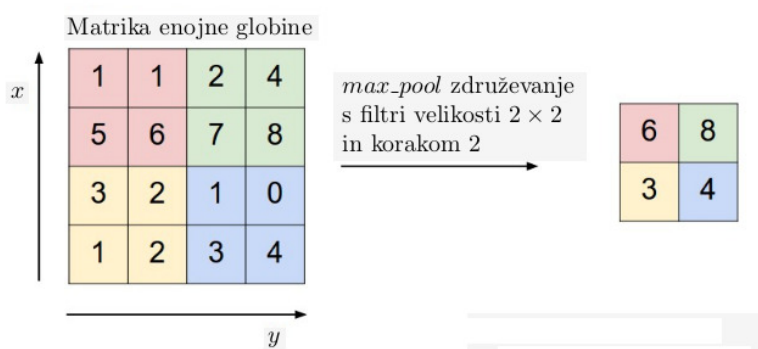
V CNN arhitekturi konvolucijskemu sloju pogosto sledi ReLU sloj. Njegova naloga je v sistem vpeljati nelinearnost s pomočjo aktivacijske funkcije (aktivacijsko funkcijo ReLU in njeno različico puščajoči ReLU smo bolj podrobno razložili v poglavju 3.1.7).

Združevalni sloj

Združevalni sloj služi za zmanjševanje števila učljivih parametrov. Manj parametrov pomeni krajši čas učenja in manjšo kompleksnost modela. S slednjim je manjša tudi verjetnost pretiranega prilagajanja podatkov (angl. overfitting). Trije pomembni parametri, ki jih moramo pri združevalnem sloju določiti, so [88]:

1. Velikost združevalnega jedra (angl. pooling kernel).
2. Korak (angl. stride).
3. Funkcija združevanja.

Slika 3.11 prikazuje operacijo združevanja. Recimo, da združevanje izvajamo nad vhodno matriko I . Velikost združevalnega jedra nam pove, nad kako velikimi podmatrikami matrike I bomo izvajali operacijo združevanja (jedro na sliki je veliko 2×2 , zato so rdeča, zelena, rumena in modra matrika velikosti 2×2). S korakom določimo, za koliko



Slika 3.11: Prikaz vpliva združevalnega jedra velikosti 2×2 [88].

se bo naše premično jedro pomikalo po matriki I – v primeru na sliki se premikamo za 2 elementa. Glede na velikost jedra, ki je prav tako 2, se izognemo prekrivanju (angl. overlapping) premičnih oken. Če si zavoljo dodatnega primera na tej sliki predstavljamo, da bi za korak izbrali vrednost 1, bi med rdečo in zeleno podmatriko dobili dodatno podmatriko z elementi $[[1, 2], [6, 7]]$, ki bi imela z rdečo podmatriko presek elementov $[[1], [6]]$ in z zeleno presek elementov $[[2], [7]]$. Hkrati bi bil končni rezultat takega združevanja matrika velikosti 3×3 . Vidimo torej, da z večanjem koraka manjšamo velikost končne matrike. Zaradi združevalnega sloja ima zato nadaljevanje arhitekture manjše število parametrov.

V primeru na sliki je uporabljena funkcija združevanja *max_pool*, ki izmed elementov izbere le tistega z največjo vrednostjo. Kot alternativo lahko uporabimo tudi kakšno drugo funkcijo (na primer izračun povprečja).

Polno-povezani sloj

V klasični konvolucijski mreži se polno-povezani sloj (angl. fully connected layer) nahaja na koncu meže – lahko gre tudi za več tovrstnih slojev. Smisel uporabe večjega števila polno-povezanih slojev je enak kakor pri uporabi večjega števila konvolucijskih slojev – naučene značilke so glede na semantično kompleksnost hierarhično urejene in izognemo se nezadostnemu prilagajanju podatkov (angl. underfitting). Izhod tega sloja gre tipično v *softmax* sloj, ki nam v fazi napovedovanja vrne končne verjetnosti za posamezno ciljno spremenljivko. Uteži iz polno-povezanih slojev (pred dokončno klasifikacijo) lahko uporabimo kot izluščevalce značilk (angl. feature extractor).

3.2.3. Uspešne arhitekture klasifikacijskih CNN

VGG

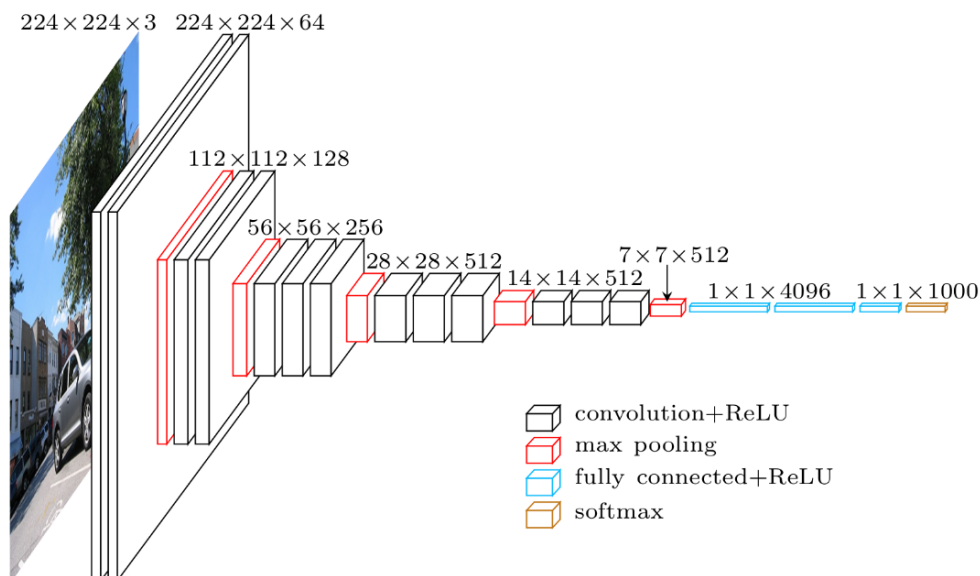
VGG [30] arhitektura je nastala po vzoru arhitekture AlexNet [29]. Njeni avtorji so preizkušali, kako globina mreže vpliva na natančnost pri razpoznavanju objektov. AlexNet je 8 nivojev globoka mreža (5 konvolucijskih slojev ter 3 polno-povezani), v delu [30] pa so ovrednotene globlje arhitekture, in sicer VGG-11, VGG-13, VGG-16 ter VGG-19, pri čemer številka v imenu predstavlja število nivojev. VGG je bila zaradi povečanja globine (in nekoliko drugačne izbire parametrov) zmagovalna arhitektura na tekmovanju ILSVRC [92] leta 2014 (kategorija klasifikacije in lokalizacije). Naloga tega tekmovanja je klasificirati slike v 1.000 različnih razredov.

VGG-16 arhitekturo sestavlja 13 konvolucijskih slojev, kakor je prikazano na sliki 3.12. Operacija združevanja *max_pool* najprej dvakrat sledi dvema zaporednima konvolucijskima slojema, nato pa trikrat sledi trem zaporednim konvolucijskim slojem. V arhitekturi so nato uporabljeni trije polno-povezani nivoji in končni *softmax* sloj. Vidimo, da je globina konvolucijskih filtrov v prvih dveh konvolucijskih slojih 64, po vsakem združevalnem sloju pa se število filtrov podvoji. Konvolucija ima relativno majhno konvolucijsko jedro velikosti 3×3 , pri čemer je korak 1. VGG-16 bomo kasneje uporabili kot osnovni gradnik arhitekture za semantično segmentacijo SegNet [31].

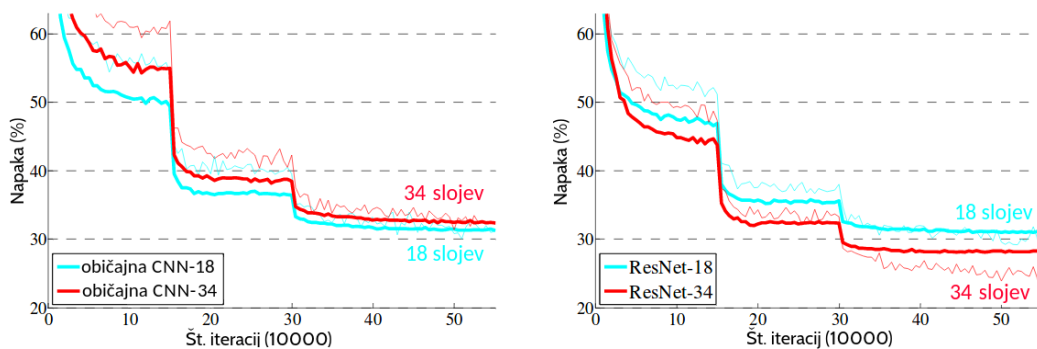
ResNet

Videli smo, da je bila arhitektura VGG v primerjavi z AlexNet uspešnejša predvsem zaradi povečane globine. Pri tem se pojavi vprašanje, ali večje natančnosti v splošnem lahko dosegamo že samo s tem, da povečujemo globino. V delu [93] so pokazali, da natančnost običajnih CNN s povečevanjem globine do neke mere narašča, potem pa začne padati, kar prikazuje prvi graf na sliki 3.13. Opazimo, da je napaka plitvejše običajne mreže z 18 sloji manjša kot njena različica z globino 34. Ta pojav imenujejo degradacija (angl. degradation) – pomembno je razumeti, da ne gre za posledico pretiranega prilagajanja podatkov. Optimizacija je otežena zaradi prevelikega števila parametrov. S povečevanjem globine običajne CNN se pogosto pojavijo tudi problemi izginjajočih/eksplodirajočih gradientov. Z uporabo ResNet se tej težavi lahko izognemo.

S pomočjo leve podslike na sliki 3.14 si predstavljajmo, da imamo dva zaporedna konvolucijska sloja *conv*₁ in *conv*₂. Vsakemu sloju sledi ReLU aktivacijska funkcija. Izhod nivoja *conv*₂ (pred drugim ReLU) označimo kot $H(x)$ – le-ta predstavlja preslikavo vhoda x . Opisani sistem imenujmo običajen blok (angl. plane block). CNN, ki jo sestavljajo običajni bloki, bomo poimenovali običajna CNN. Poleg običajnih blokov so v običajni



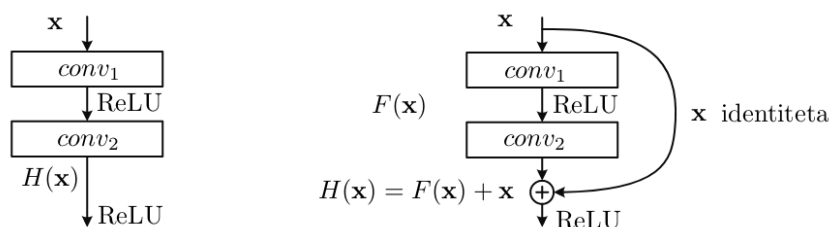
Slika 3.12: Arhitektura VGG-16 [30].



Slika 3.13: Primerjava natančnosti običajne CNN arhitekture in ResNet arhitekture glede na spreminjanje globine [93]. Tanjše črte prikazujejo napako na učni množici, debelejšje pa napako na validacijski množici. Če običajno CNN (pri čemer številčna oznaka predstavlja globino mreže) poglobljamo od neke globine naprej, se njena natančnost zmanjša (prvi graf). Z uporabo ResNet arhitekture pa se problemu degradacije lahko izognemo (drugi graf), kar omogoči, da s poglobljanjem ResNet arhitekture izboljšamo natančnost mreže.

CNN tipično prisotni tudi združevalni sloji in *softmax* sloj. Ob učenju običajne CNN torej optimiziramo več običajnih blokov, cilj optimizacije pa je narediti preslikavo $H(x)$ čim bolj ustrezno glede na zadan problem. Če je običajna CNN, ki jo poskušamo učiti pregloboka, se zaradi prevelike globine pojavi problem degradacije.

Izkaže se [93], da se nam običajne bloke splača dopolniti tako, da vsakemu izmed njih dodamo preskočno povezavo (angl. skip connection) identitete, kakor prikazuje desna podsvila na sliki 3.14. Na tak način dopolnjen običajen blok imenujemo blok preostankov (angl. residual block).



Slika 3.14: Prikaz razlike med običajnim blokom in blokom preostankov [93].

Hipotezi avtorjev sta naslednji:

1. Če je identiteta x pri bloku preostankov optimalna (optimalen $H(x)$ je x), je preslikavo $F(x)$ bolj enostavno narediti nevtralnno (jo npr. izenačiti z 0), kakor pa optimizirati uteži v slojih $conv_1$ in $conv_2$ pri običajnem bloku. Običajen blok namreč nima preskočne povezave identitete x , hkrati pa je ob neničelnih utežeh blokov $conv_1$ in $conv_2$ prisotna nelinearna aktivacijska funkcija.
2. Če je mapiranje $H(x) = F(x) + x$ pri bloku preostankov blizu optimalnega (identiteta opiše večino potrebne informacije), nam ni potrebno veliko spreminjati uteži na nivojih $conv_1$ in $conv_2$.

ResNet se v praksi dobro izogne problemu degradacije. V fazi vzratnega popravljanja napake se gradienti lahko širijo preko preskočnih povezav in dolgih povezav ostankov, kar omogoča učinkovito učenje. S povečevanjem globine se izboljšuje tudi natančnost modela, dokler se zaradi globine mreže model ne začne preveč prilegati podatkom. ResNet arhitektura bo kasneje naš gradnik za segmentacijsko arhitekturo RefineNet.

3.2.4. Uspešne arhitekture segmentacijskih CNN

Celostno-konvolucijske mreže

V prejšnjem poglavju smo predstavili arhitekture konvolucijskih mrež, ki so se izkazale za uspešne pri reševanju klasifikacijskih problemov. Končna napoved takih mrež je ponavadi enodimenzionalen vektor, iz katerega lahko razberemo, kakšne so verjetnosti, da je na vhodni sliki korespondenčni objekt. Obstajajo pa tudi problemi, ki sodijo v t.i. razred gostih napovednih nalog (angl. dense prediction tasks) [94], pri katerih moramo verjetnost razreda napovedati za vsak posamezen slikovni element na sliki. V ta razred sodijo na primer naslednji problemi:

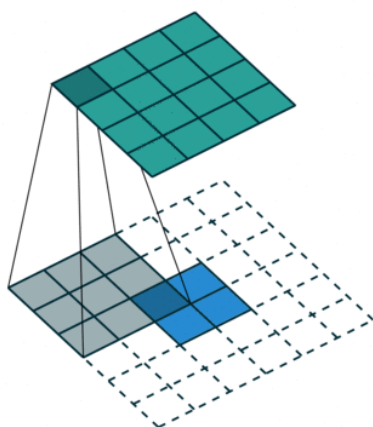
1. Semantična segmentacija (angl. semantic segmentation).
2. Eno-očesna ocena globine (angl. monocular depth estimation).
3. Napovedovanje robov (angl. boundary prediction).
4. Restavriranje slik.
5. Povečevanje ločljivosti slik (angl. super-resolution).

Avtorji dela [94] so pokazali, da lahko obstoječe klasifikacijske arhitekture spremenimo in dopolnimo tako, da postanejo uspešne tudi pri tovrstnih nalogah. Avtorji arhitekturo imenujejo celostno-konvolucijska mreža (angl. fully convolutional network, FCN), pomembne ideje dela pa so:

1. Konvolucionalizacija (angl. convolutionalization) oziroma zamenjava polno-povezanih (angl. fully-connected layers) s konvolucijskimi sloji. Na polno-povezan sloj lahko gledamo tudi kot na konvolucijo z enim jedrom, ki je tako veliko, da prekrije vse vhodne regije. Prva slabost polno-povezanega sloja je, da na izhodu dobimo podatke brez prostorske dimenzije. Druga slabost je, da moramo zaradi vnaprej določene dimenzije na vhod podajati slike fiksne velikosti (ne moremo na primer spreminjati velikosti vhodnih slik). Konvolucijski sloj odpravi obe pomanjkljivosti. V FCN končne odločitve sprejemajo torej filtri konvolucijskega sloja.
2. Dekonvolucijski sloj (angl. deconvolutional layer) natančneje transponirani konvolucijski sloj (angl. transposed convolutional layer) ali delno koračna konvolucija (angl. fractionally strided convolution). Po zadnjem konvolucijskem sloju pri običajni CNN (ki je v mreži zaradi konvolucionalizacije zamenjal polno-povezani sloj), imamo v zadnji mapi značilik zaradi združevalnih slojev zelo malo prostorske informacije. FCN mora na izhod vrniti sliko, ki ima enako velikost kakor vhodna slika. Zato si želimo imeti mehanizem, s katerim bi mapam značilik, ki imajo na

tem nivoju nizko ločljivost in veliko globino, posredovali prostorsko informacijo, na podlagi katere se bo ločljivost izhodne slike povečala. Slika 3.15 prikazuje, na kakšen način ta sloj poveča vhodno sliko, ki je označena z modro barvo. Transponirani konvolucijski sloj (kakor navadni konvolucijski sloji) vsebuje filtre, katerih uteži v sistemu optimiziramo. Mreža se torej sama nauči, na kakšen način izhodno sliko povečati (angl. upsampling).

3. Preskočna arhitektura (angl. skip architecture). Glavna ideja je, da posredujemo informacijo iz slojev, ki opisujejo nizkonivojske značilke v sloje, ki opisujejo visokonivojske abstraktne značilke (angl. high level features). V našem primeru potrebujemo informacijo iz finih nizkonivojskih slojev zato, da v procesu povečevanja slike lahko rekonstruiramo prostorsko informacijo (kje se objekt nahaja). Na ta način združimo informacijo, ki opisuje, kje se objekt nahaja (bolj plitvi sloji) ter kaj objekt je (globlji sloji).

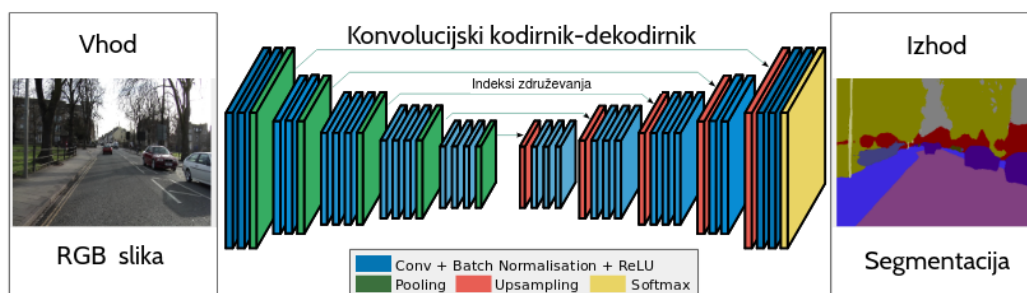


Slika 3.15: Prikaz transponiranega konvolucijskega sloja, pri čemer je modri del originalna slika, sivi kvadrat pa sprejemno polje. Jedro transponirane konvolucije je veliko 3×3 , velikost koraka je 1.

SegNet

SegNet segmentacijska arhitektura je v osnovi zgrajena iz dveh delov: kodirnika in dekodirnika. Kodirnik je enak arhitekturi VGG-16, s to razliko, da so odstranjeni polno-povezani sloji. To je koristno, ker lahko po principu prenešenega učenja (angl. transfer learning) inicializiramo uteži, ki so se na drugih domenah izkazale za uspešne. Dekodirnik si lahko interpretiramo kot dodaten, obrnjen VGG-16, brez polno-povezanih slojev, ki mu dodamo

softmax sloj. V arhitekturi imamo tudi preskočne povezave, ki gredo iz združevalnih slojev v korespondenčne sloje za zvišanje kvalitete slike (angl. upsampling). Te preskočne povezave nosijo informacijo o tem, kateri indeks je imel pri združevanju v jedru največjo vrednost (v združevalnem sloju je namreč uporabljena funkcija *max_pool*). Ta informacija omogoča kvalitetno prostorsko rekonstrukcijo (ustrezne filtre se mreža nauči sama), zaradi česar je segmentacija natančna. Mreža ima enake velikosti vhodne in izhodne slike (kakor smo videli pri FCN). Za konvolucijskimi sloji je za preprečevanje pretiranega prilaganja podatkov in hitrejšo konvergenco modela uporabljena paketna normalizacija (angl. batch normalization). Izhod zadnjega *softmax* sloja je slika s K kanali, kjer je K enak številu razredov. Element (i, j) , ki pripada posameznemu kanalu, predstavlja verjetnost, da slikovni element (i, j) na vhodni sliki pripada razredu, ki ga opisuje ta kanal.



Slika 3.16: Arhitektura SegNet je osnovana na arhitekturi VGG-16 [31].

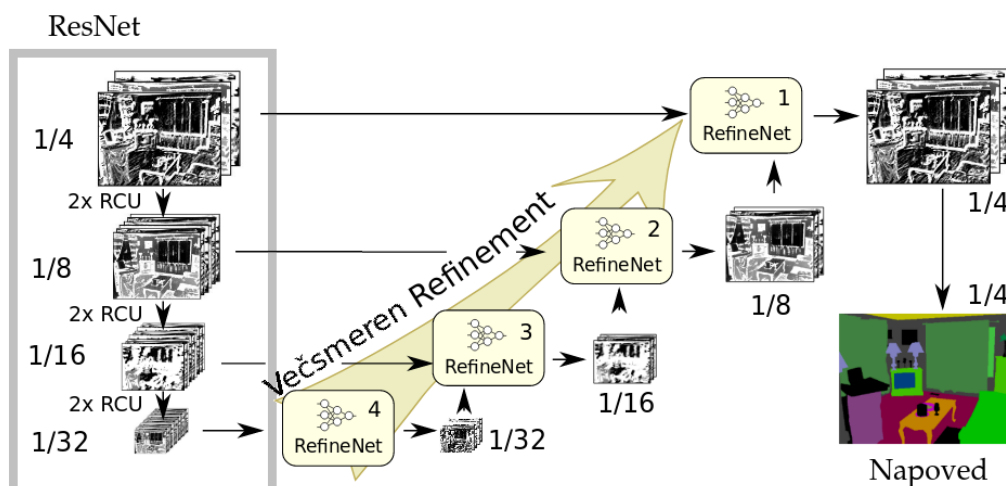
RefineNet

Arhitekturo RefineNet, ki je prikazana na sliki 3.17, v grobem sestavljata dva visokonivojska gradnika [95]:

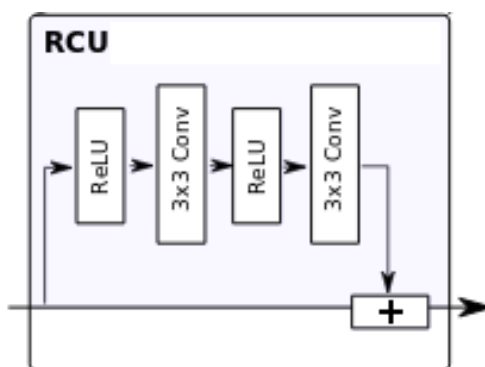
1. ResNet arhitektura s konvolucijskimi enotami preostanka (angl. residual convolution unit, RCU). Vidimo, da se (vertikalno gledano) širina in višina map značilno zmanjšujeta za polovico, medtem ko se število kanalov dvakrat poveča. Vsaka izhodna mapa značilnik služi kot eden izmed vhodov v soležni RefineNet blok.
2. RefineNet blok. Prvi vhod vanj je soležna mapa značilnik, drugi pa izhod RefineNet bloka iz nižjega nivoja (če le-ta obstaja). Potrebno je omeniti tudi, da je v delu [95] RCU opisan kot del RefineNet bloka. Ker v tem delu želimo poudariti vsebovanost ResNet arhitekture v RefineNet arhitekturi, RCU ne bo opisan kot del RefineNet bloka.

S to zasnovo RefineNet zagotavlja, da so v drugem delu mreže zastopane značilke iz vseh nivojev (angl. multi level features). RCU, katerega struktura ustreza definiciji bloka

preostankov, je prikan na sliki 3.18. V RCU imamo na daljši poti dve ponovitvi ReLU aktivacijskega sloja, ki mu sledi konvolucija z velikostjo jedra 3×3 .

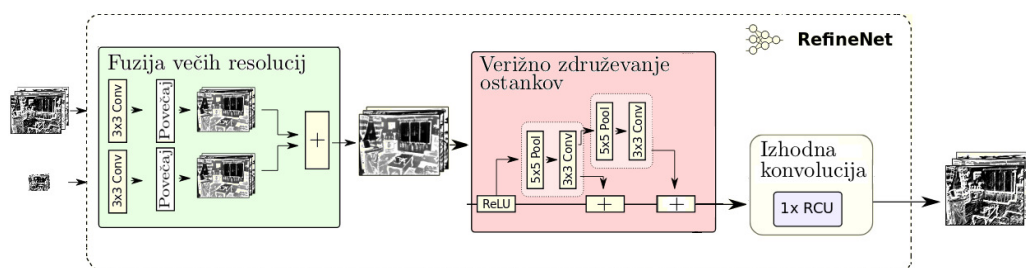


Slika 3.17: Prikaz arhitekture RefineNet [95]. Kodirnik je sestavljen iz mreže ResNet, pri čemer je vsak nivo map značilk eden izmed vhodov v istoležni RefineNet blok (drugi vhod v RefineNet blok je izhod RefineNet bloka nižjega nivoja).



Slika 3.18: Prikaz RCU, ki ustreza definiciji bloka ostanka [95].

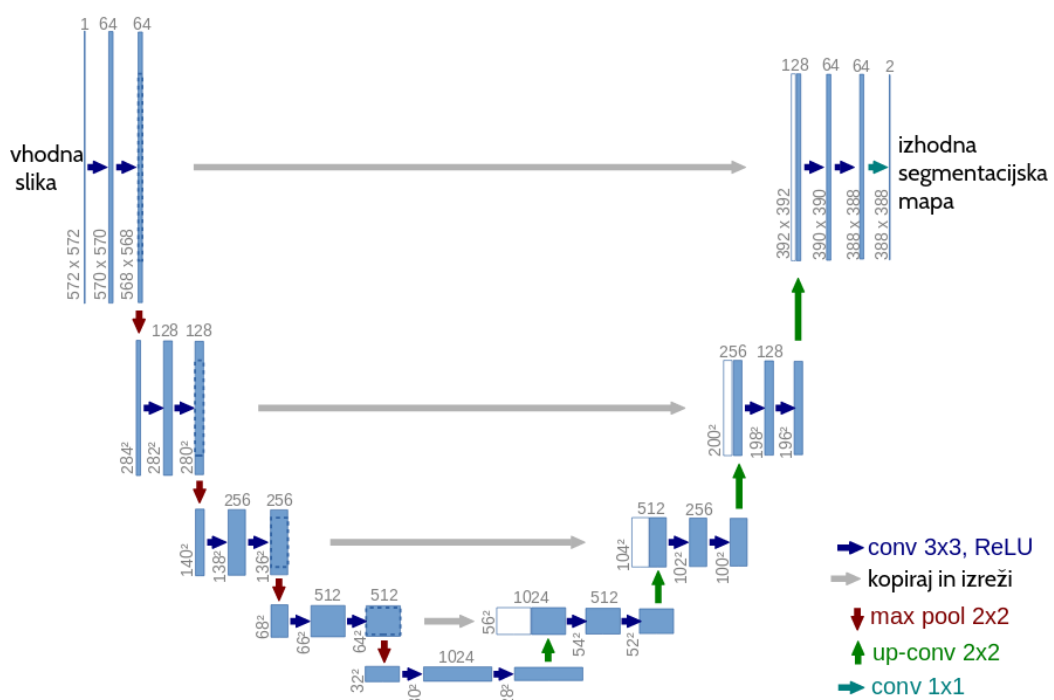
Na sliki 3.19 je natančneje predstavljena zgradba RefineNet bloka. Kakor smo že omenili, sta vhoda (razen na najnižjem nivoju) tipično dva: mapa značilk višjega nivoja ter izhod RefineNet bloka nižjega nivoja. Vsak vhod posebej gre najprej skozi 3×3 konvolucijo, nato sta obe sliki povečani na enako velikost ter sešteti skupaj. Izhod je nato poslan v t.i. verižno združevanje ostankov (angl. chained residual pooling), za katerega avtorji trdijo, da zajame kontekst ozadja.



Slika 3.19: Natančnejši prikaz RefineNet bloka [95]. Na vhod dobimo mape značilk iz istoležnega nivoja in iz RefineNet bloka nižjega nivoja. Oba vhoda povečamo na isto velikost in ju seštejemo. Verižno združevanje ostankov zajame kontekst ozadja.

UNet

UNet je arhitektura [96], ki se je za uspešno izkazala pri reševanju biomedicinskih problemov (segmentacija različnih vrst celic in določanje mej med dotikajočimi se celicami). Njena prednost je, da se dobrih modelov lahko nauči že iz relativno majhnega števila učnih primerov – pri mnogih biomedicinskih problemih namreč nimamo na voljo veliko učnih primerov, za učenje te mreže pa pogosto zadostuje že močna obogatitev učnih podatkov. Arhitektura mreže (prikazana na sliki 3.20) za rekonstrukcijo prostorske informacije uporablja transponirano konvolucijo, informacijo pa v globlje sloje posreduje preko operacije kopiraj in izreži (angl. copy and crop).



Slika 3.20: Arhitektura segmentacijske mreže UNet [96].

Poglavje 4

Podatkovne zbirke

4.1. MASD

MASD je podatkovna zbirka, ki je bila izdana v sklopu tekmovanj SS(ER)BC [20, 44]. Pri izdelavi zbirke so 82 ljudem slikali obe očesi (164 različnih očes). Vsako oko je slikano tako, da gleda v 4 različne smeri: naravnost, levo, desno in navzgor. Pogleda navzdol ni zato, ker ljudje ob pogledu navzol avtomatsko zapremo zgornjo veko, kar prekrije regijo interesa (beločnico in šarenico). Za vsako oko pri izbrani smeri pogleda, so zbrani štirje vzorci, kar na posamezno oko nanese 16 slik (4 smeri pogleda \times 4 vzorci). Celotna zbirka je tako sestavljena iz 2.624 RGB slik, ki so zajete v visoki ločljivosti 300 pik na palec (angl. dots per inch, DPI) pri dimenzijah 7.500×5.000 . Za slikanje je bil uporabljen NIKON D 800. Javno je dostopen le manjši del zbirke (učna množica 120 slik), večinski del (testna množica) pa ni odprtega tipa. V podatkovni zbirki sta zastopana oba spola. Vsa očesa so rjave barve, ostale barve niso zastopane. Šarenica v mnogih primerih ni kvalitetno zajeta (npr. odsev bliskavice, neizostreno), saj je zbirka fokusirana na zajem beločnice. V šarenici se na mnogih slikah odsevajo trepalnice, zato pogosto ni razločno vidno, ali gre za trepalnice ali za njihov odsev. To oteži anotiranje šarenice in obrvi (ki smo jih potrebovali za razvoj algoritmov za semantično segmentacijo očesa). Mnogo beločnic je močno zamegljenih in žile v beločnici niso jasno vidne, kar poveča težavnost prepoznavanja na podlagi beločnice. Učna množica vsebuje tudi 120 anotacij beločnice na nivoju slikovnih elementov (angl. per-pixel annotations). Na večini slik ni zajetih obrvi (periokularen del je manj zastopan), kar omeji uporabnost podatkovne zbirke za razvoj algoritmov, ki prepoznavajo na podlagi periokularnega dela. Težava te zbirke je tudi to, da imena datotek mnogokrat niso konsistentna, prav tako pa niso opisna (iz imena se ne

da ugotoviti, za katero oko gre in v katero smer gleda), kar oteži delo z zbirko. Ugotovili smo tudi, da so slike nekaterih očes v zbirki podvojene oz. pripadajo drugim osebam, kar zahteva dodatno prečiščevanje zbirke.

4.2. UBIRIS.v2

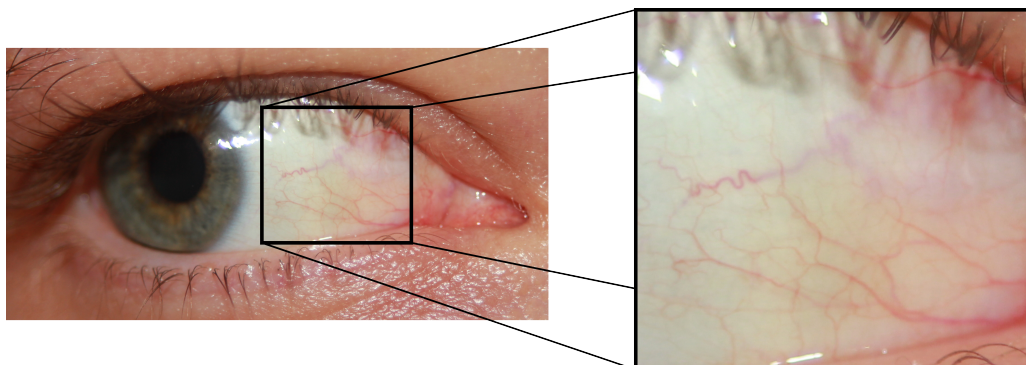
Podatkovna zbirka UBIRIS.v2 [48] predstavlja naslednjo verzijo zbirke UBIRIS.v1 [47]. Prva verzija nas v tem delu ne bo zanimala. UBIRIS.v2 vsebuje 11.102 slik, ki pripadajo 261 različnim osebam (v zbirki je torej 522 različnih očes). Za vsako oko je zbranih 15 ali 30 primerov slik, pri čemer so zastopani pogledi naravnost, levo in desno (pogleda navzgor ni). Slike so bile zajete z aparatom Canon EOS 5D, in so v zbirki dostopne v ločljivosti 400×300 slikovnih elementov. Zbirka je javno dostopna. Oba spola sta približno enakomerno zastopana (54.4% moških in 45.6% žensk). Zastopane so vse glavne skupine očesnih barv. K zbirki ni priloženih nobenih anotacij na nivoju slikovnih elementov. Zbirka je prav tako uporabna za razpoznavanje na podlagi periokularnega dela, saj primerki slik vsebujejo tako slike, kjer je vidna širša obočesna regija (vključno z obrvmi in delom nosa), kot tudi slike, kjer je viden le ožji del obočesne regije (vidne so gube in pore). Poimenovanje datotek je konsistentno, vendar se iz njega ne da sklepati, ali gre za levo ali desno oko, oziroma za kateri pogled gre.

4.3. SBVPI

V tem podpoglavju predstavimo novo podatkovno zbirko, ki smo jo poimenovali po njeni namembnosti – namenjena je raziskavam razpoznavanja na podlagi žil iz beločnice, periokularne regije in šarenice (angl. Sclera Blood Vessels, Periocular and Iris, SBVPI). Podatkovna zbirka je za raziskovalne namene javno dostopna na povezavi <http://sclera.fri.uni-lj.si>.

4.3.1. Motivacija

Na voljo so številne podatkovne zbirke za očesno biometrijo [12], vendar je pri večini zbirk glaven poudarek na raziskovanju prepoznavanja na podlagi šarenice. Te zbirke bi bile potencialno lahko uporabljene tudi za raziskovanje prepoznavanja na podlagi beločnice, vendar je problem ta, da so pogosto zajete v NIR spektru, v katerem se večina informacije o beločnici izgubi. Poleg tega je veliko obstoječih zbirk nastalo tako, da avtorji pri zajemu



Slika 4.1: Primer slike iz zbirke SBVPI s približno regijo, ki prikazuje vaskularne vzorce v beločnici.

fotografij niso bili pozorni na vaskularne strukture, zaradi česar fotografije pogosto ne vsebujejo regije interesa (angl. region of interest, ROI), ki so potrebne za razpoznavo na podlagi beločnice. Da bi naslovili ta problem, so bile nedavno izdelane podatkovne zbirke, kot je na primer MASD [20], vendar le-te niso javno dostopne. Da bi olajšali raziskovanje prepoznavanja na podlagi žil v beločnici, v nadaljevanju predstavljamo javno odprto podatkovno zbirko SBVPI, ki je ena izmed redkih zbirk, ki je posebej namenjena raziskovanju segmentacije beločnice in prepoznavanja na podlagi beločnice. Poleg tega so v zbirki tudi slike, ki so namenjene raziskovanju obočesnega dela.

4.3.2. Opis podatkovne zbirke SBVPI

Podatkovna zbirka SBVPI je sestavljena iz dveh delov. Prvi del je zbirka slik, ki so prvotno namenjene raziskovanju vaskularne biometrije, drugi del pa je namenjen predvsem raziskavam periokularne biometrije.

Vaskularni del SBVPI

Vaskularni del zbirke, ki je prvotno namenjen raziskovanju beločnice in ožje definirane periokularnega dela, vsebuje 1.959 RGB slik, ki pripadajo 55 različnim osebam. Slike so bile zajete z zrcalno refleksno kamero (Canon EOS 60D) pri najvišji možni ločljivosti in kvaliteti. Da bi maksimizirali zajem podrobnosti (majhne žilice, kožne pore in gube), smo uporabljali tudi makro leče. Z opisanimi postopki zajema smo dosegli, da so vzorci žil v regiji beločnice na fotografijah jasno vidni, kar prikazuje slika 4.1.

Med zajemom slik je bila kamera pozicionirana v spremenljivi razdalji med 20-40 centimetrov od osebe. Da bi zagotovili dodatno različnost med vzorci, ki pripadajo isti identiteti, smo pred zajemom vsake fotografije kamero naključno premaknili iz prejšnje

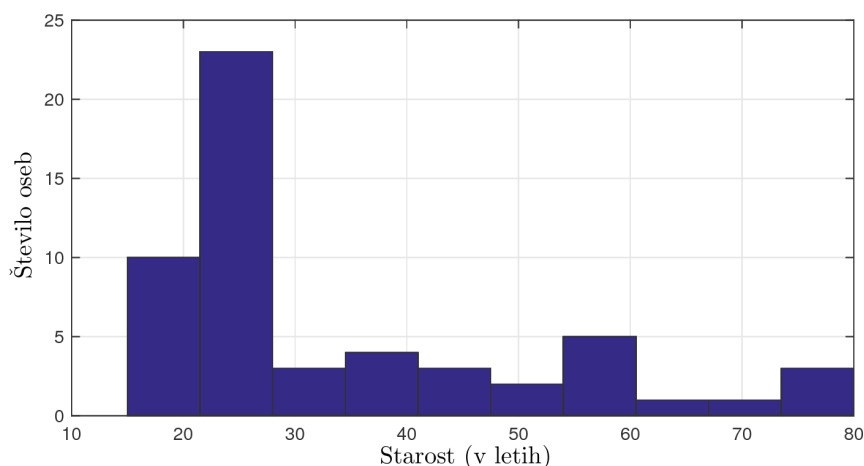


Slika 4.2: Primeri slik iz podatkovne zbirke SBVPI. Podatkovna zbirka vsebuje visoko kvalitetne primerke z jasno vidnimi strukturami žil v beločnici. Za vsako osebo v zbirki je bilo zbranih vsaj 32 slik, ki vsebujejo obe očesi s štirimi različnimi smermi pogleda (naravnost/levo/desno/navzgor). Zgornji dve vrstici prikazujeta 8 slik osebe moškega spola, spodnji dve vrstici pa 8 slik osebe ženskega spola.

pozicije za približno 5-30 centimetrov levo/desno/navzgor/navzdol. Med spreminjanjem pozicije kamere je tudi oseba, ki smo jo fotografirali, nekoliko spremenila pozicijo svoje veke in smer pogleda. S takim postopkom zajemanja smo zagotovili, da se posamezni primerki slik istega očesa, ki gleda v isto smer, vedno razlikujejo med seboj. Znano je namreč, da majhne spremembe v smeri pogleda osebe povzročijo kompleksne nelinearne deformacije vidnih vaskularnih struktur v beločnici [38]. Naša želja je bila izdelati zbirko, primerno za razvoj algoritmov, ki bi bili odporni na tovrstne deformacije.

Primeri slik so včasih zajemali neželene dele obraza (npr. obrvi, dele nosu, itd.). Zato smo odvečne dele odstranili z ročnim obrezovanjem, pri čemer je bilo pri obrezovanju uporabljeno fiksno razmere (angl. aspect ratio). S tem smo zagotovili, da je na končnih slikah vsebovan le ozek del periokularne regije, kar je prikazano tudi na sliki 4.2.

Pri procesu zajemanja slik smo se zgedovali po procesu zajemanja, ki je bil uporabljen pri izdelavi zbirke MASD [20]. Vsako osebo smo prosili, naj pogleda v eno izmed smeri naravnost/levo/desno/navzgor. Za vsako izmed teh smeri smo zajeli eno sliko in jo shranili v podatkovno zbirko. Ta proces smo štirikrat ponovili, posebej za levo in desno oko. Tako je nastalo za vsako osebo minimalno 32 različnih slik (4 ponovitve \times 4 smeri pogleda \times 2 očesi). Nekatere osebe so bile pri posameznem pogledu slikane tudi več kot štirikrat.



Slika 4.3: Starostna porazdelitev oseb, ki so prisotne v podatkovni zbirki. Osebe so stare med 15 in 80 let, pri čemer je večina oseb stara manj kot 35 let.

Med zajemanjem smo sproti preverjali, če so slike dovolj kvalitetne (če so dovolj ostre in nezamegljene), pri čemer smo slike zavrgli, če niso zadostovale subjektivnim kriterijem. Če smo sliko zavrgli, je bila zajeta druga slika. Osebe, ki nosijo očala, smo prosili, če jih lahko snamejo, medtem ko je bilo nošenje leč med fotografiranjem dovoljeno. Prav tako smo si prizadevali, da bi bilo na zajetih slikah prisotno čim manj odbojev svetlobe, ki jih povzroči bliskavica.

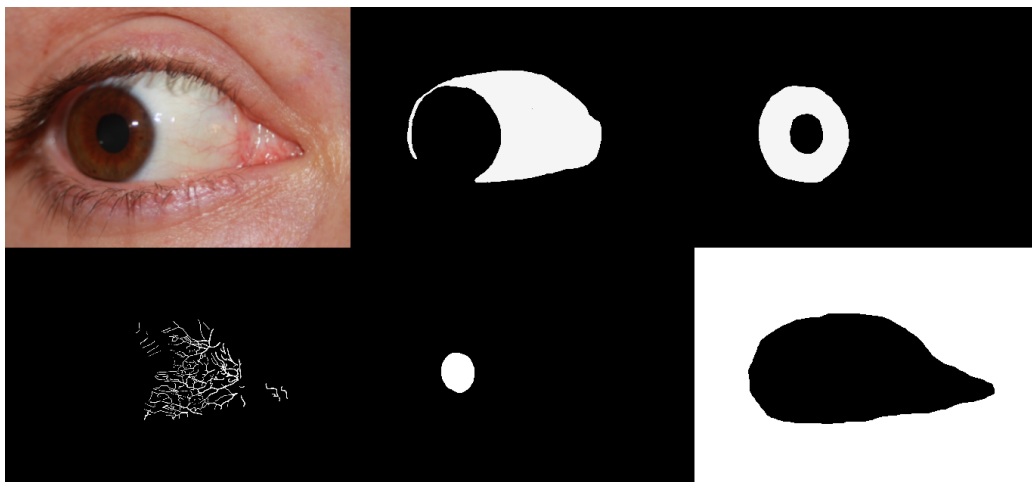
Končna podatkovna zbirka je uravnotežena glede na spol in vsebuje 29 žensk ter 26 moških oseb, pri čemer vse osebe pripadajo beli oziroma kavkazijski rasi. Starost oseb je v razponu med 15 in 80 let, pri čemer je večina oseb stara manj kot 35 let, kar prikazuje tudi starostni histogram na sliki 4.3. SBVPI se od MASD razlikuje tudi po tem, da so v SBVPI prisotne različne barve oči, kar predstavlja dodaten kriterij različnosti v zbirki. Povzetek karakteristik zbirke SBVPI je predstavljen v tabeli 4.1.

Podatkovna zbirka je anotirana z identiteto (55 različnih identitet), spolom (ženski, moški), razredom očesa (levo ali desno oko) in oznako smeri pogleda (naravnost, levo, desno, navzgor), pri čemer so vse anotacije dostopne za vsako izmed 1.959 slik. Poleg tega so dodane ročno izdelane anotacije posameznih delov očesa na nivoju slikovnih elementov (angl. per-pixel annotations). Zbirka vsebuje 1.959 slik, na katerih je anotirana regija beločnice. Poleg tega je trenutno na voljo približno 200 (to število se bo v prihodnosti povečalo) anotacij šarenice in zenice ter 130 anotacij vaskularnih struktur. Primeri anotacij so prikazani na sliki 4.4. Segmentacijske maske so bile narejene ročno s pomočjo programa GIMP. Vsaka slika v SBVPI ima svoj GIMP projekt, v katerem je več slojev. Vsak posamezen sloj vsebuje eno vrsto anotacije (beločnico, zenico, šarenico itd.) na

nivoju slikovnih elementov. Dostopne anotacije naredijo našo podatkovno zbirko ustrezno za odprto raziskovanje algoritmov razpoznavanja na podlagi beločnice in raziskovanje segmentacije beločnice, česar ostale podobne zbirke ne omogočajo. Predvsem so ročno izdelane anotacije vaskularnih struktur unikaten prispevek SBVPI zbirke.

Tabela 4.1: Glavne karakteristike zbirke SBVPI.

Karakteristika	Opis
Fotoapararat	DSLR fotoapararat, Canon EOS 60D + makro leče
Število slik	1.959
Število identitet	55
Število slik na identiteto	Vsaj 32 različnih slik
Velikost slik	1.700×3.000 slikovnih elementov
Dostopne anotacije	Identiteta, spol, smer pogleda, segmentacijske mape

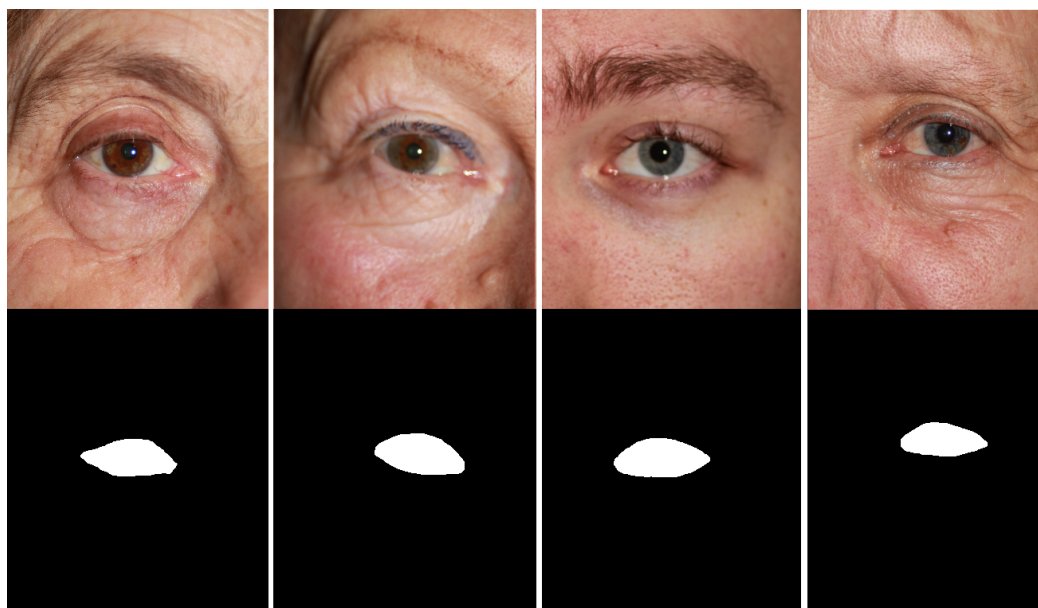


Slika 4.4: Primeri anotacij na nivoju slikovnih elementov, ki so del podatkovne zbirke SBVPI. Na sliki so prikazane naslednje slike (od leve proti desni, od zgoraj navzdol): slika očesa, ročna anotacija beločnice, anotacija šarenice, anotacija vaskularnih struktur, anotacija zenice ter anotacija periokularne regije.

Periokularni del SBVPI

Drugi del zbirke SBVPI, ki je namenjen predvsem raziskovanju širše definirane periokularnega dela, vsebuje istih 55 identitet kakor prvi del zbirke, slike pa so zajete po

drugačnem protokolu. Na slikah je v drugem delu zbirke prikazan širši periokularni del, h kateremu poleg očesa spadajo tudi obrvi, del lica ter del nosu. Za vsako identiteto smo zbrali po 8 slik, pri čemer 4 slike prikazujejo periokularni del levega očesa in 4 slike periokularni del desnega očesa. Drugi del zbirke skupaj torej sestavlja 440 slik, za vsako sliko pa je dostopna tudi ročna anotacija periokularnega dela. Primeri slik s korespondenčnimi anotacijami iz periokularnega dela SBVPI so prikazani na sliki 4.5.



Slika 4.5: Prikaz periokularnega dela zbirke SBVPI, ki je namenjen raziskovanju širše definicije periokularnega dela. V zgornji vrstici vidimo slike periokularnega dela, ki pripadajo različnim identitetam, v spodnji vrstici pa so korespondenčne ročno izdelane anotacije, pri čemer bela barva označuje očesni del, črna barva pa periokularno regijo.

4.4. Primerjava podatkovnih zbirk

V tabeli 4.2 primerjamo opisane podatkovne zbirke. Vidimo, da je prednost SBVPI predvsem veliko število raznovrstnih anotacij (levo/desno oko, smer pogleda in različnih delov očesa na nivoju slikovnih elementov). SBVPI ima pred MASD to prednost, da je javno dostopna in vsebuje očesa, ki zastopajo vse barvne skupine. V primerjavi z drugimi zbirkami SBVPI vsebuje manjše število identitet in slik, vendar se temu problemu izognemo z razširitvijo zbirke, s katero število identitet in slik povečamo za faktor 2 (kar bo opisano v nadaljevanju dela). Pri UBIRIS.v2 imamo sicer na voljo več različnih smeri pogledov,

vendar se iz vrstnega reda fotografij (oziroma imen datotek) ne da sklepati, za katero smer pogleda gre na sliki.

Tabela 4.2: Primerjava zbirk MASD, UBIRIS.v2 in SBVPI.

Karakteristika	MASD	UBIRIS.v2	SBVPI
v celoti javno dostopna	ne	da	da
št. identitet	82	261	55
št. slik	2.624	11.102	2.399
zastopanost barv oči	samo rjave	vse barvne skupine	vse barvne skupine
različne smeri pogleda	naravnost/ levo/ desno/ navzgor (nekonsistenten protokol)	naravnost/ levo/ desno (nekonsistenten protokol)	naravnost/ levo/ desno/ navzgor (konsistenten protokol)
	anotacije očesa (levo/desno)	ne	da
	anotacije smeri pogleda	ne	da
	anotacije beločnice	da	ne
anotacije šarenice	ne	ne	da
anotacije zenice	ne	ne	da
anotacije periokluarnega dela	ne	ne	da

Poglavje 5

Prepoznavanje na podlagi beločnice

5.1. Uvod

Zaradi naraščajočih potreb po varnih avtentikacijskih sistemih, forenzičnih aplikacijah ter nadzornih (angl. surveillance) sistemih je v interesu raziskovalnih skupin in privatnih podjetij izboljšati trenutno stanje tehnologije. V primerjavi z drugimi biometričnimi modalnostmi ima očesna biometrija številne prednosti. Te so: brezstično zajemanje podatkov, visoka natančnost prepoznavanja in sprejemanje tovrstne tehnologije s strani uporabnikov. Na področju očesne biometrije je trenutno najbolj razširjeno prepoznavanje na podlagi šarenice. Trenutne raziskave se ukvarjajo z uporabo dodatnih očesnih karakteristik, ki bi lahko dopolnjevale prepoznavanje na podlagi značilk, pridobljenih iz šarenice. Z uporabo dodatnih karakteristik bi bilo mogoče zmanjšati število prevar in s tem povečati zanesljivost sistemov.

Modalnost, ki ima v tem kontekstu velik potencial, je struktura žil v beločnici. Regija beločnice vsebuje veliko vaskularnih struktur, za katere velja, da so pri vsakem posamezniku različne in so relativno stabilne med procesom staranja [4]. Zato so vaskularne strukture primerne, da jih uporabimo v razpoznavalne namene [20, 22]. Vzorci vaskularnih struktur imajo tudi druge zaželjene lastnosti [41], ki jih naredijo primerne za razpoznavalne sisteme. Strukture so na primer še vedno vidne, tudi če je oko pordelo, hkrati pa ostanejo nespremenjene v primeru, da bi vsiljivec nosil umetne leče, ki bi spremenile barvo šarenice. Kljub vsem naštetim prednostim je raziskovanje te modalnosti še vedno v zgod-

njih fazah in naslovljenih mora biti še veliko izzivov, preden bo tovrstna tehnologija lahko uporabljena v komercialnih sistemih.

V tem poglavju naslovimo nekatere izmed teh izzivov in predstavimo nov pristop za razpoznavanje na podlagi beločnice, česar se lotimo z uporabo globokega učenja in konvolucijskih nevronske mreže. Najprej predstavimo novo tehniko za segmentacijo vaskularnih struktur, ki se nahajajo v beločnici, z uporabo kaskade SegNet mrež [31]. Predlagana tehnika nam omogoča natančno segmentacijo žil na vhodnih slikah, kar olajša luščenje značilke (angl. feature extraction) in računanje ujemanja (angl. matching) v kasnejših fazah cevovoda. Nato ovrednotimo dele segmentacijskega cevovoda s konkurenčnimi segmentacijskimi modeli, ki bazirajo na metodah globokega učenja, kot sta na primer RefineNet [95] in UNet [96]. Naredimo tudi primerjavo s segmentacijskim modelom, ki temelji na tradicionalnih metodah strojnega učenja. Naši eksperimenti pokažejo, da predlagan segmentacijski pristop (osnovan na SegNet arhitekturi) zagotavlja konkurenčne segmentacijske zmogljivosti in predstavlja učinkovito rešitev za dotični problem.

Za razpoznavalni del cevovoda predstavimo CNN model, ki smo ga poimenovali ScleraNET in je bil razvit v sklopu našega dela. ScleraNet učinkovito izlušči diskriminantne značilke iz segmentiranih žil beločnice. Da bi zagotovili, da so izluščene značilke odporne na morebitne spremembe smeri pogleda, mrežo ScleraNet učimo z večkriterijskim učenjem (angl. multi-task learning, MTL). Mreža se (poleg učenja identitete) uči napovedati tudi za katero smer pogleda gre. Predlagan model ScleraNET ovrednotimo z identifikacijskimi in verifikacijskimi eksperimenti. S ScleraNET dobimo vzpodbudne rezultate, ki jih primerjamo s konkurenčnim tradicionalnim algoritmom SIFT.

5.2. Metode

V tem podpoglavju je predstavljen naš pristop za prepoznavanje na podlagi beločnice. Začeli bomo z visokonivojskim pregledom cevovoda, kasneje pa bomo bolj podrobno opisali posamezne komponente.

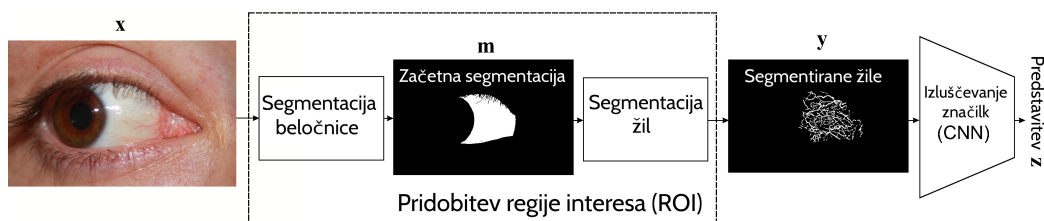
5.2.1. Pregled cevovoda

Visokonivojski pregled cevovoda za razpoznavanje na podlagi beločnice, ki je opisan v tem poglavju, je predstavljen na sliki 5.1. Cevovod sestavljata dva glavna gradnika: *i*) kaskada konvolucijskih nevronske mreže, ki pridobi regijo interesa (konkretno žil), ter *ii*) konvolucijska nevronska mreža ScleraNET, ki izračuna numerično predstavitev slike.

Kaskada konvolucijskih nevronske mreže na vhod sprejme RGB sliko očesa in iz nje pridobi verjetnostno mapo vaskularnih struktur beločnice. To stori v dveh korakih: *i*)

mreža za segmentacijo beločnice najprej napove, kje se nahaja beločnica, nato pa *ii*) ansambel nevronske mreže iz regije beločnice (ki je bila pridobljena v prejšnjem koraku) napove, kje se nahajajo žile.

Drugi gradnik cevovoda (ScleraNET) na vhod sprejme verjetnostne mape žil in vrne diskriminativno predstavitev, ki je lahko uporabljena za izračun ujemanja. V naslednjem podglavju sta podrobno opisana oba dela našega cevovoda.



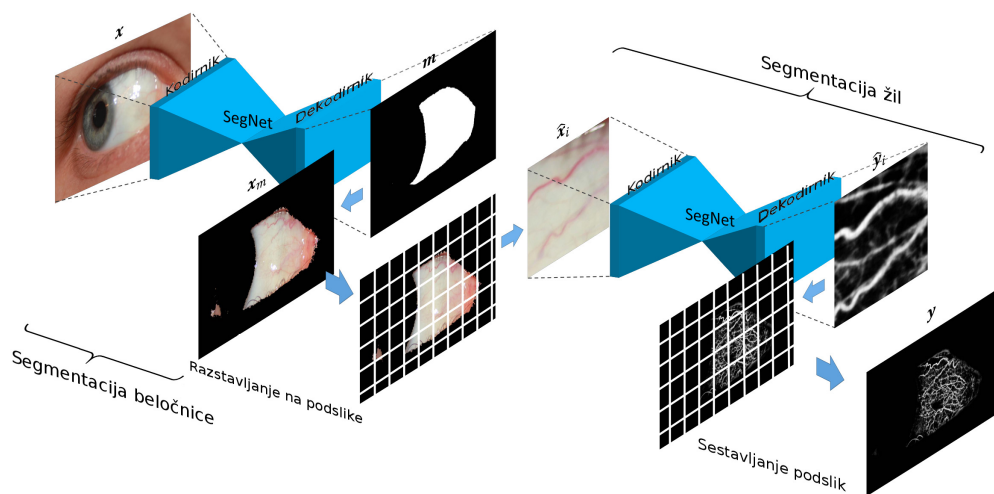
Slika 5.1: Prikaz predlaganega bločnega diagrama, ki opisuje algoritem za razpoznavanje na podlagi beločnice. Najprej so z uporabo dvo-fazne segmentacijske procedure segmentirane vaskularne strukture v predelu beločnice. Verjetnostna mapa vaskularnih struktur je nato podana CNN modelu (imenovanemu ScleraNET), ki izlušči diskriminativne značilke. Slednje so uporabljene za izračun ujemanja.

5.2.2. Pridobitev regije interesa

Eden izmed ključnih korakov vsakega biometričnega sistema je pridobitev regije interesa. V sistemu, ki razpozna na podlagi beločnice, je pridobitev regije interesa enaka procesu, s katerim segmentiramo vaskularne strukture iz vhodne slike. Vsak posameznik ima unikatno vaskularno strukturo in jo zato lahko uporabimo za namene razpoznavanja [38]. Kakor je bilo izpostavljeno v prejšnjem podglavju, moramo na vhodni sliki najprej poiskati vaskularno strukturo, kar naredimo z dvo-fazno proceduro, ki je narejena iz kaskade konvolucijskih nevronske mreže.

5.2.3. Dvo-fazna segmentacijska procedura

Kaskada konvolucijskih nevronske mreže, ki smo jo v našem cevovodu uporabili za pridobitev regije interesa, je prikazana na sliki 5.2. Sestavljena je iz dveh segmentacijskih CNN modelov, kjer prvi ustvari binarno masko, ki označuje regijo beločnice na vhodni sliki, drugi model pa segmentira vaskularne strukture, ki so prisotne v beločnici. Oba segmentacijska modela sta osnovana na predstavljeni arhitekturi SegNet [31].



Slika 5.2: Ilustracija dvo-fazne segmentacijske procedure. V prvi fazi se s SegNet modelom ustvari binarna maska, ki opiše pozicijo beločnice na vhodni sliki. Masko uporabimo za prekrivanje delov vhodne slike, ki niso relevantni za drugo fazo, v kateri je cilj z uporabo drugega SegNet modela poiskati vaskularne strukture. Da bi bili zmožni zajeti majhne detajle v vaskularnih strukturah, je druga faza implementirana kot razstavljanje slike na podslike ter sestavljanje rezultatov, ki jih dobimo iz posameznih podslik (angl. patch-wise approach).

Če vhodno RGB sliko očesa označimo kot \mathbf{x} in binarno masko beločnice (vrne jo prvi SegNet model) označimo kot \mathbf{m} , potem lahko prvo segmentacijsko fazo formalno zapišemo kot:

$$\mathbf{m} = f_{\theta_1}(\mathbf{x}), \quad (5.1)$$

kjer f_{θ_1} označuje mapiranje vhodne slike \mathbf{x} v segmentacijski rezultat \mathbf{m} , ki ga vrne prvi CNN model, in θ_1 predstavlja parametre, ki jih je potrebno naučiti v fazi učenja mreže.

Ko je beločnica segmentirana, maskiramo vhodno sliko \mathbf{x} z pridobljeno segmentacijsko masko \mathbf{m} . S tem izločimo vse slikovne elemente, ki ne pripadajo beločnici, rezultat maskiranja pa gre v nadaljnje procesiranje. Operacijo maskiranja matematično lahko opišemo kot:

$$\mathbf{x}_m = \mathbf{x} \odot \mathbf{m}, \quad (5.2)$$

kjer \odot predstavlja Hadamardov produkt. Maskirana vhodna slika \mathbf{x}_m je nato uporabljena kot osnova za drugo segmentacijsko fazo.

Ker vaskulaturu beločnice sestavljajo večje in manjše (tanjše) žile, v drugi segmentacijski fazi uporabimo način s sestavljanem in razstavljanjem slike (angl. patch-wise approach). Ta pristop nam omogoča, da model zazna velike in hkrati manjše strukture žil,

ki bi bile pri uporabi celostnega pristopa izgubljene ali prezrte zaradi slabega kontrasta ali pa neopazne. Maskirano vhodno sliko \mathbf{x}_m razdelimo na M enako velikih, neprekrivajočih se delov $\{\hat{\mathbf{x}}_i\}_{i=1}^M$ in vsak del vstavimo v drugi segmentacijski model f_{θ_2} , ki napove, kje se nahaja vaskularna struktura $\hat{\mathbf{y}}_i$:

$$\hat{\mathbf{y}}_i = f_{\theta_2}(\hat{\mathbf{x}}_i), \quad \text{for } i = 1, \dots, M. \quad (5.3)$$

V tej enačbi θ_2 označuje parametre drugega SegNet modela, ki mora biti prav tako naučen na nekih učnih podatkih.

Končna verjetnostna mapa vaskularnih struktur \mathbf{y} je pridobljena iz ponovne združitve posameznih podslik $\hat{\mathbf{y}}_i$ z mozaičenjem. Pomembno je opaziti, da nam za razliko od prvega segmentacijskega modela, ki vrne binarno masko, drugi segmentacijski model vrne mapo verjetnosti. Ugotovili smo namreč, da za razpoznavalne namene verjetnostna mapa deluje bolje, kakor če bi uporabili binarno masko vaskularne strukture.

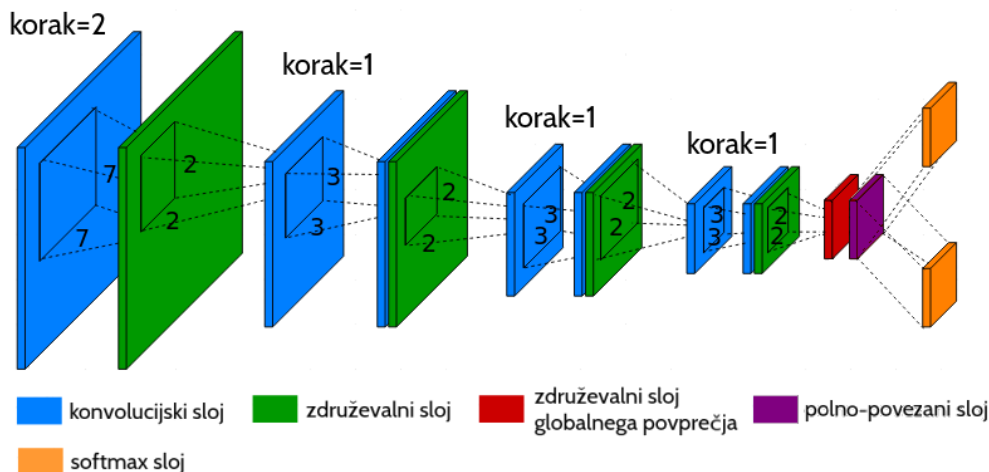
Da bi bil segmentacijski algoritem za pridobitev verjetnostne mape vaskularnih struktur čim bolj odporen, v času izvajanja uporabljamo tehniko obogatitve podatkov (angl. data augmentation) na sledeč način. Iz maskirane slike \mathbf{x}_m izdelamo več njenih verzij, pri čemer posamezna verzija nastane tako, da \mathbf{x}_m rotiramo za naključen kot, jo naključno zrcalimo čez x in/ali y os ter jo naključno premaknemo (angl. shift). Vsaka izmed teh novo nastalih verzij je razdeljena na M podslik ter neodvisno poslana v proces segmentacije vaskularnih struktur. Potem so napovedi podslik M posamezne verzije sestavljene nazaj v celotno napoved dotične verzije. Iz vsake verzije obogatene \mathbf{x}_m dobimo torej eno verjetnostno mapo. Ker je vsaka verzija \mathbf{x}_m nastala z naključnimi premiki, rotacijami in zrcaljenji, napovedi verzij niso semantično poravnane. Poravnamo jih tako, da nad verjetnostno mapo posamezne verzije izvedemo inverzne operacije, ki so bile uporabljene za nastanek te verzije. S tem dobimo več semantično poravnanih napovedi za celotno sliko \mathbf{x}_m , kot končno rešitev pa vzamemo povprečje teh napovedi.

Učenje modela in generiranje izhoda

Da bi lahko naučili oba SegNet modela f_{θ_1} in f_{θ_2} , ki ju potrebujemo za našo segmentacijsko proceduro, za učenje uporabimo kategorično navzkrižno entropijo (angl. categorical cross-entropy). Ko je model enkrat naučen, le-ta za vsak slikovni element vrne verjetnostno porazdelitev razredov C ciljnih spremenljivk. V naših primerih sta ciljni spremenljivki bodisi *beločnica in drugo* bodisi *vaskularne strukture in drugo*, torej je v obeh primerih $C = 2$. Za vsak slikovni element vhodne slike $s = [x, y]^T$ model vrne distribucijo $\mathbf{p}_s = [p_{C_1}, p_{C_2}]^T \in \mathbb{R}^{C \times 1}$, kjer p_{C_i} označuje verjetnost, da slikovni element na lokaciji s pripada i -temu ciljnemu razredu C_i , pri čemer velja $\sum_{i=1}^C p_{C_i} = 1$ [97]. Model torej vrne dve verjetnostni mapi, ki sta invertirani verziji ena druge, saj $p_{C_1} = 1 - p_{C_2}$.

Ko potrebujemo binarne segmentacijske rezultate (npr. v primeru pridobivanja maske

regije beločnice \mathbf{m}), je mapa verjetnosti binarizirana z uporabo v naprej definirane praga Δ . Ta prag lahko nastavimo na vrednost med 0 in 1, upravljanje pa je definirano tako, da vse vrednosti, ki so manjše ali enake Δ , postavimo na 0, vse ostale pa na 1.



Slika 5.3: Pregled arhitekture ScleraNET.

5.2.4. ScleraNET za razpoznavanje

V drugem delu našega cevovoda se zanašamo na CNN arhitekturo, ki smo jo poimenovali ScleraNET in služi kot izluščevalec značilk (angl. feature extractor). Arhitektura je bila razvita v sklopu tega dela. ScleraNET značilke pridobi iz verjetnostnih map vaskularnih struktur. Pomembno je poudariti, da so razpoznavalne tehnike na podlagi vidnih vaskularnih struktur zelo občutljive na spremembe smeri pogleda. Smer pogleda namreč vpliva na količino vidne vaskulature in posledično na zmogljivosti končnega razpoznavalnega pristopa. Naš cilj je ScleraNET naučiti, da bo iz verjetnostnih map vaskularnih struktur zmožen izluščiti diskriminativne značilke, ki bodo maksimalno odporne na spremembe smeri pogleda. Zato mrežo učimo z večkriterijskim učenjem (angl. multi-task learning), pri katerem sta dva kriterija, na podlagi katerih se optimizirajo parametri modela, in sicer: *i*) identiteta ter *ii*) smer pogleda. V naslednjem podpoglavju se nahaja podroben opis arhitekture ScleraNET, procedure učenja ter opis, kako mrežo uporabimo kot izluščevalca značilk.

Tabela 5.1: Povzetek zgradbe arhitekture ScleraNET.

Št.	Tip sloja	Št. filtrov	Opis
1.	konvolucijski	128	7×7 (korak velikosti 2)
2.	združevalni		2×2
3.	konvolucijski	128	3×3 (korak velikosti 1)
4.	konvolucijski	128	3×3 (korak velikosti 1)
5.	združevalni		2×2
6.	konvolucijski	256	3×3 (korak velikosti 1)
7.	konvolucijski	256	3×3 (korak velikosti 1)
8.	združevalni		2×2
9.	konvolucijski	512	3×3 (korak velikosti 1)
10.	konvolucijski	512	3×3 (korak velikosti 1)
11.	združevalni		2×2
12.	združevalni globalnega povprečja		
13.	polno-povezani	512	
14.	softmax ($2 \times$)		Večkriterijsko učenje

Arhitektura mreže ScleraNET

Arhitektura ScleraNET je zgrajena po vzoru nedavno predlaganih CNN modelov, ki uspešno rešujejo probleme razpoznavanja in vključuje arhitekturne odločitve, ki jih uporabljajo pri mrežah AlexNet [29] in VGG [30]. Kot je prikazano na sliki 5.3 in povzeto v tabeli 5.1, arhitekturo sestavlja 7 konvolucijskih slojev (z ReLU aktivacijo), med katerimi se nahaja več združevalnih slojev. Sledijo združevalni sloj globalnega povprečja (angl. global average pooling layer), polno-povezani sloj in na koncu dva *softmax* sloja.

Da lahko zajamemo dovolj prostorskega konteksta in zmanjšamo velikost map značilk v nadaljevanju cevovoda, prvi konvolucijski sloj uporablja 128 relativno velikih filtrov velikosti 7×7 s korakom 2. Sledi združevalni sloj, ki za faktor 2 še nadalje zmanjša velikost map značilk po obeh dimenzijah. Nato so uporabljeni trije zaporedni bloki, sestavljeni iz dveh konvolucijskih in enega združevalnega sloja. Zaradi združevalnih slojev je prostorska dimenzija map značilk po vsakem bloku prepolovljena. Da bi modelu zagotovili zadostno moč za kvalitetno reprezentacijo map značilk, po vsakem združevalnem sloju podvojimo število filtrov v konvolucijskih slojih. Izhod tretjega bloka je podan v združevalni sloj globalnega povprečja (angl. global average pooling) in nadalje v 512-dimenzijski polno-povezani sloj. Na koncu je polno-povezani sloj povezan na dva *softmax* sloja, nad katerima

sta v fazi učenja mreže definirani funkciji izgube – prva izračuna izgubo napovedi identitete, druga pa izgubo napovedi smeri pogleda. *Softmax* sloja nista uporabljena v času, ko ScleraNET uporabljamo za luščenje značilk.

Učenje modela

Za oba *softmax* sloja, ki se nahajata na koncu ScleraNET arhitekture, definiramo izgubo prečne entropije (angl. cross-entropy loss). Prva izguba prečne entropije L_1 kaznuje napake, ki nastanejo, ko model napačno napove identiteto, druga izguba prečne entropije L_2 pa kaznuje napake, ki nastanejo pri napačni napovedi smeri pogleda. Končna napaka pri večkriterijskem učenju je definirana kot:

$$L_{skupna} = L_1 + L_2. \quad (5.4)$$

Da bi se naučili parametrov mreže ScleraNET, minimiziramo kombinirano izgubo na podlagi podanih učnih podatkov. Pri tem za izračun skupne napake obema členoma izgube damo enako utež.

Intuicija v ozadju večkriterijskega učenja je, da model hkrati naučimo reprezentacij značilk, ki so uporabne za obe nalogi, torej razpoznavanja identitete in smeri pogleda. Poleg tega si lahko izgubo, ki meri napake ob napačni klasifikaciji pogleda, interpretiramo kot regularizator za proces ugotavljanja identitete [98]. Dodaten člen do neke mere prispeva k učenju bolj diskriminativnih opisnikov značilk, ki znajo ločiti med različnimi smermi pogledov.

Napovedovanje identitete z mrežo ScleraNET

Ko je ScleraNET model naučen, ga naredimo uporabnega za identitete, ki jih model še ni videl tako, da iz arhitekture odstranimo oba *softmax* sloja in polno-povezani sloj. 512-dimenzionalni izhod združevalnega sloja globalnega povprečja nato uporabimo kot vektor značilk, ki opisuje na vhod podano vaskularno strukturo.

Če ponovno označimo verjetnostne mape vaskularnih struktur (ki jih pridobimo ob koncu dvo-fazne segmentacijske procedure) kot \mathbf{y} , potem lahko izračun predstavitve značilk opišemo kot:

$$\mathbf{z} = g_{\theta}(\mathbf{y}), \quad (5.5)$$

kjer g_{θ} označuje mapiranje vaskularne strukture \mathbf{y} v predstavitve značilk \mathbf{z} , ki jih vrne model ScleraNET, pri čemer so θ parametri modela.

5.3. Eksperimenti in rezultati

V tem podpoglavju ovrednotimo naš cevovod za razpoznavanje na podlagi beločnice. Podpoglavje začnemo z opisom eksperimentalnega protokola in uporabljenih metrik uspešnosti (angl. performance metrics). Nato opišemo učne postopke za vse dele našega cevovoda, na koncu pa predstavimo rezultate in jih prediskutiramo. Da bi zagotovili ponovljivost naših rezultatov, vse modele, podatke, anotacije in eksperimentalne skripte javno objavljamo na povezavi <http://sclera.fri.uni-lj.si>.

5.3.1. Metrike uspešnosti

Končna uspešnost našega razpoznavalnega cevovoda je v veliki meri odvisna od uspešnosti delovanja segmentacijskega dela cevovoda, ki ga uporabljamo za pridobitev vaskularnih struktur. Prav tako je zelo odvisna od diskriminativne moči reprezentacij značilk, ki so izluščene iz segmentirane vaskulature. Zato v podpoglavju z opisom eksperimentov naredimo ločene eksperimente za segmentacijski del cevovoda ter ločene eksperimente za del, kjer merimo uspešnost izluščevanja značilk. V nadaljevanju opišemo metrike uspešnosti, ki jih uporabljamo v obeh delih eksperimentov.

Metrike uspešnosti za segmentacijske eksperimente: Uspešnost segmentacijskega modela merimo s standardnimi metrikami uspešnosti, kot so *natančnost* (angl. precision), *priklic* (angl. recall) in *F₁-ocena* (angl. F₁-score), ki so definirane na sledeč način [97, 99, 100]:

$$\text{natančnost} = \frac{TP}{TP + FP}, \quad (5.6)$$

$$\text{priklic} = \frac{TP}{TP + FN}, \quad (5.7)$$

$$F_1\text{-ocena} = 2 \cdot \frac{\text{natančnost} \cdot \text{priklic}}{\text{natančnost} + \text{priklic}}, \quad (5.8)$$

kjer *TP* označuje število pravilno pozitivnih (angl. true positive) slikovnih elementov, *FP* število napačno pozitivnih (angl. false positive) slikovnih elementov in *FN* število napačno negativnih (angl. false negative) slikovnih elementov.

Med zgoraj naštetimi merami natančnost meri razmerje med pravilno segmentiranimi slikovnimi elementi in številom vseh pravilnih slikovnih elementov ciljnega razreda (regija beločnice). Natančnost opiše, koliko segmentiranih slikovnih elementov je v resnici

relevantnih. Priklic meri delež pravilno segmentiranih slikovnih elementov, v primerjavi s številom vseh slikovnih elementov, ki pripadajo ciljnemu razredu, s čimer opiše informacijo o tem, koliko relevantnih slikovnih elementov je najdenih/segmentiranih. Natančnost in priklic sta tipično odvisna – eno mero lahko povečamo na račun zmanjševanja druge mere (in obratno), kar lahko reguliramo s spreminjanjem praga. Če je potrebno na preprost način primerjati dva segmentacijska modela, lahko uporabimo F_1 -oceno. Le-ta na preprost način združi natančnost in priklic v eno samo metriko, ki jo kot dodatno metriko uporabljamo tudi v tem delu [97].

Opazimo, da ob fiksaciji segmentacijskega praga Δ za ovrednotenje segmentacijskega rezultata dobimo fiksni vrednosti za natančnost in priklic. Za predstavitev kompromisa (angl. trade-off) med merama pa lahko uporabimo vizualizacijo v obliki tako imenovane krivulje natančnost/priklic (angl. precision/recall curve). Za izris te krivulje moramo pri vseh spreminjajočih se segmentacijskih pragovih Δ izračunati obe meri. Tako vizualiziran kompromis prikaže bolj celostno podobo uspešnosti segmentacijskega modela.

Metrike uspešnosti za eksperimente razpoznavanja: Tukaj merimo uspešnost cevovodnega dela, ki izlušči značilke (in razpoznavna) z metrikami za ovrednotenje verifikacije ter identifikacije.

Poročamo o uspešnosti verifikacije s standardno oceno napačnega sprejemanja (angl. false acceptance rate, FAR) ter standardno oceno napačne zavrnitve (angl. false rejection error, FRR). FAR meri napako nelegitimnih poskusov verifikacije, FRR pa meri napako legitimnih verifikacijskih poskusov. Obe napaki, FAR in FRR, sta odvisni od izbranega praga (podobno kot smo videli pri natančnosti in priklicu). Če izberemo prag, ki proizvede nižji FAR, to povzroči zvišanje FRR in obratno – če izberemo prag, ki proizvede nižji FRR, to povzroči zvišanje FAR. Obe napaki sta omejeni na vrednost med 0 in 1. Splošno sprejeta praksa biometričnih raziskav je, da se namesto FRR ocene predstavi verifikacijsko oceno (angl. verification rate, VER), ki je definirana kot $1 - \text{FRR}$ [101, 102, 103, 104]. VER lahko označimo tudi kot standardno oceno pravega sprejemanja (angl. true acceptance rate, TAR).

Da lahko prikažemo celosten kompromis med FAR in FRR (oziroma VER), izrišemo krivuljo karakteristike delovanja sprejemnika (angl. receiver operating characteristic, ROC), za prikaz katere moramo izračunati vse možne vrednosti odločevalnega praga. V naših eksperimentih nato poročamo o nekaterih točkah, npr. uspešnost verifikacije pri $\text{FAR} = 1\%$ ($\text{VER}@1\text{FAR}$), $\text{FAR} = 10\%$ ($\text{VER}@10\text{FAR}$) in pri tako imenovani oceni enake napake (angl. equal error rate, EER), ki je korespondenčna točka na ROC, kjer sta FAR in FRR enaka. Poleg tega za vse razpoznavalne eksperimente predstavimo tudi ploščino pod ROC krivuljo (angl. area under the curve, AUC), kar je pogosta mera za ocenjevanje natančnosti biometričnega sistema, ki ga uporabljamo v verifikacijski postavitvi.

Za poročanje o uspešnosti identifikacijskih eksperimentov uporabimo razpoznavalni

oceni Rank-1 in Rank-5, ki merita delež identifikacijskih poskusov, kjer je razpoznavalni sistem pravilni identiteti pripisal najvišje ujemanje, oziroma je pravilna identiteta ena izmed petih najvišjih ujemanj. Nato enako kot v delu [105] predstavimo še dobljeno ploščino pod normalizirano kumulativno krivuljo ujemanja (angl. area under the normalized cumulative match score curve, AUCMC).

Naši rezultati so grafično predstavljeni tudi v obliki kumulativne krivulje ujemanja (angl. cumulative match score curve, CMC), ki izriše ocene razpoznavanja (angl. recognition rate) za vse razpoznavalne range. Na ta način prikazuje bolj celostno stanje uspešnosti naših eksperimentov identifikacije.

5.3.2. Eksperimentalni protokol in učni detajli

Eksperimente izvajamo na zbirki SBVPI, ki smo jo opisali v podpoglavju 4.3 z uporabo ločenih eksperimentalnih protokolov za segmentacijski in razpoznavalni del našega cevovoda. V nadaljevanju je predstavljen protokol in podrobnosti učnih postopkov.

Eksperimenti segmentacije

Segmentacijski del našega cevovoda sestavljata dve komponenti. Prvi na vhodni sliki segmentira regijo beločnice, drugi pa zazna, kje v beločnici se nahajajo vaskularne strukture.

Segmentacija beločnice: Za učenje in testiranje segmentacijskega modela prve komponente našega cevovoda razdelimo SBVPI del podatkovne zbirke, ki je namenjen preučevanju beločnice, v dve disjunktni množici:

- *Učna množica* je sestavljena iz 1.223 slik očes s korespondenčnimi maskami beločnice. Množica je razdeljena še v dve podmnožici. V prvi podmnožici je 1.039 slik, ki so uporabljene za učenje parametrov modela. V drugi množici je 184 slik, ki jih med učenjem uporabljamo kot validacijsko množico, zato da na njej lahko sproti spremljamo generalizacijo modela oz. ugotavljamo, ali je že prišlo do pretiranega prilagajanja podatkov (angl. over-fitting).
- *Testna množica* je sestavljena iz 736 slik očes s korespondenčnimi maskami beločnice. Ta množica je v eksperimentih uporabljena za končno ovrednotenje delovanja in izračun metrik uspešnosti.

Da bi se izognili pretiranemu prilagajanju podatkov, učno množico (1.039 slik) obogatimo (angl. augment). To naredimo tako, da iz vsake slike naredimo 40 novih slik, in sicer z naključnimi operacijami zrcaljenja, prirezovanja, megljenja z Gaussovim jedrom (angl. Gaussian blurring), spremembami osvetlitve ter z afinimi transformacijami, kot sta na primer skaliranje in rotacije (do $\pm 35^\circ$).

Učenje SegNet modela za prvo segmentacijsko fazo (segmentacijo beločnice) je opravljeno na grafični kartici GTX 1080 Ti z 11GB RAM-a.

Za eksperimente uporabljamo javno dostopno Caffe implementacijo¹. Vhodne slike so za fazo učenja pomanjšane na fiksno velikost 360×480 slikovnih elementov. Uteži modela so naučene z uporabo SGD in inicializirane z metodo iz dela [106]. Stopnja učenja (angl. learning rate) je nastavljena na 0,001, padanje uteži (angl. weight decay) na 0,0005, moment na 0,9 ter velikost paketov (angl. batch size) na 4. Model konvergira po 26.000 iteracijah.

Segmentacija vaskulature: Druga komponenta našega cevovoda zahteva anotacije vaskularnih struktur na nivoju slikovnih elementov za obe proceduri: učenje in testiranje. Zbirka SBVPI vsebuje skupaj 130 anotiranih slik. Te podatke (slike s korespondenčnimi anotacijami) ponovno razdelimo v dve disjunktni množici:

- *Učna množica* vsebuje 98 slik, ki so razdeljene na podslike velikosti 360×480 slikovnih elementov. Na ta način iz njih dobimo 788 podslik, od katerih jih 630 naključno izberemo za učenje parametrov modela. Preostalih 158 podslik v fazi učenja uporabimo kot našo validacijsko množico. Da bi se izognili pretiranemu prilagajanju podatkov, ponovno obogatimo učne podslike, tako da iz vsake, z naključnimi operacijami (rotacije, obrezovanje, manipulacije barv), izdelamo 40 novih slik.
- *Testna množica* je sestavljena iz 32 slik. Ker so testne slike ponovno procesirane na nivoju podslik, poročamo o rezultatih, pridobljenih nad celotnimi slikami in ne nad vmesno predstavitevijo s podslikami.

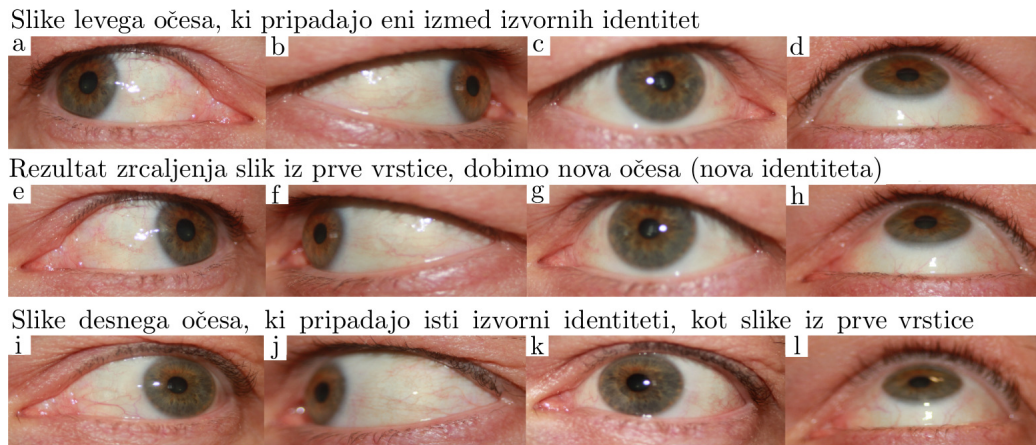
Za učenje segmentacijskega modela vaskularnih struktur uporabimo enake nastavitve, kot je opisano pri modelu za segmentacijo beločnice.

Razširitev podatkovne zbirke SBVPI

Kot smo omenili v podpoglavju 4.3, osnovna zbirka SBVPI vsebuje 55 različnih identitet, za eksperimente razpoznavanja pa bomo iz njih ustvarili več identitet. To prvotno število identitet bomo zaradi bolj jasne razlage v nadaljevanju poimenovali *izvirne identitete*. Predstavili bomo dva načina, kako smo za namene naših eksperimentov razpoznavanja število končnih identitet razširili, in sicer *i*) zrcaljenje očes čez os y ter *ii*) upoštevanje levega in desnega očesa kot samostojni identiteti.

Zbirko SBVPI lahko razširimo tako, da slike očes posamezne izvorne identitete prezrcalimo čez y os in prezrcaljene slike označimo kot novo identiteto. Iz vsake prvotne identitete lahko torej naredimo eno novo. Na ta način iz 55 izvornih identitet naredimo 55 novih identitet, kar je skupaj 110 identitet. Razširjeno množico 110 identitet (55 izvornih identitet + 55 prezrcaljenih identitet = 110 identitet) bomo z avtorji bolj jasne razlage

¹Dostopno na: <https://github.com/alexgkendall/caffe-segnet>



Slika 5.4: Prikaz načina razširitve osnovne zbirke SBVPI. V prvi vrstici so slike levega očesa izvorne identitete. Z zrcaljenjem ustvarimo slike nove razširjene identitete, ki so prikazane v drugi vrstici – vidimo, da zdaj le-te predstavljajo desno oko, levi in desni pogled pa se glede na prvo vrstico zamenjata. Tretja vrstica predstavlja dejanska desna očesa izvorne identitete. Opazimo, da se vaskularne strukture druge in tretje vrstice močno razlikujejo, zaradi česar je z zrcaljenjem možno razširiti zbirko.

poimenovali kot *razpoznavalne identitete*, saj jih bomo skušali napovedati v naših eksperimentih razpoznavanja. Za novo nastalo oko sta po zrcaljenju čez y os značilni dve stvari: *i*) oko postane nasprotno-ležno oko (če je na izvorni sliki prikazano levo oko, bo na prezrcaljeni sliki prikazano desno oko) ter *ii*) na sliki prikazan pogled levo (oziroma desno) ob zrcaljenju postane pogled desno (oziroma levo).

Zakaj je opisana razširitev SBVPI možna, je grafično prikazano na sliki 5.4. V prvi vrstici so prikazane vse 4 različne smeri pogledov levega očesa, ki pripadajo izvorni identiteti. Novo identiteto naredimo tako, da izvorne slike prezrcalimo čez y os, kar prikazuje druga vrstica (če zrcalimo sliko a , dobimo sliko e , z zrcaljenjem b dobimo f itd.). Če je prva vrstica prikazovala levo oko, druga vrstica predstavlja desno oko. Opazimo, da se pogled levo/desno ob zrcaljenju zamenja, pri pogledih naravnost/gor pa ostane enak. Zdaj nas zanima še, koliko se prezrcaljena očesa (ki zdaj predstavljajo desno oko nove identitete) razlikujejo od dejanskega desnega očesa izvorne identitete. Zato v tretji vrstici prikažemo vse 4 poglede desnega očesa izvorne identitete, pri čemer se vrstni red predstavljenih pogledov tretje vrstice ujema z vrstnim redom pogledov v drugi vrstici (e in i obe predstavljata desno oko s pogledom desno, f in j s pogledom levo itd.). Vidimo, da se vaskularne strukture med očesi izvorne in novoustvarjene identitete pri istoležnem očesu in pogledu povsem razlikujejo, kar je razlog, da na opisan način lahko razširimo podatkovno zbirko.

Vaskularna struktura beločnice je epigenetska karakteristika, za katero je značilno, da je pri vsaki osebi edinstvene oblike. Razlikuje se celo med levim in desnim očesom istega posameznika. Za namene učenja naših modelov je zato levo in desno oko razpoznavalne identitete smiselno definirati kot ločeni identiteti. Ta metodologija je značilna za epigenetske biometrične značilnosti in je bila pogosto uporabljena v literaturi, npr. [107, 108]. Na ta način iz 110 razpoznavalnih identitet dobimo 220 identitet, ki jih bomo poimenovali kot *učne identitete*, saj jih bomo uporabljali v fazi učenja modelov. Tako po dve učni identiteti (levo in desno oko iste osebe), pri katerih si zapomnimo, iz katere razpoznavalne identitete smo jih dobili, kažeta na isto razpoznavalno identiteto.

Če torej povzamemo, smo iz 55 izvornih identitet naredili 220 učnih identitet, ki kažejo na 110 razpoznavalnih identitet.

Ekspirimenti razpoznavanja

Za razpoznavalne eksperimente razdelimo razširjeno zbirko SBVPI 220-ih učnih identitet na učno in testno množico, ki je disjunktna glede na identitete. Pri razdelitvi upoštevamo tudi, da se izvorna in z zrcaljenjem ustvarjena identiteta pojavita v isti množici. Učno množico sestavlja 2.074 slik, ki pripadajo 120 različnim identitetam. Te slike so naprej razdeljene v dejanske učne podatke (potrebne za učenje parametrov modela) ter validacijsko množico (uporabljena za sprotno spremljanje učenja) v razmerju 70% in 30%. Preostalih 1.630 slik, ki pripadajo 100 identitetam, je uporabljenih za končno testiranje.

Za učenje parametrov ponovno uporabljamo grafično kartico GTX 1080 Ti. Naš ScleraNET model je implementiran v okolju Keras in inicializiran po metodi, ki je opisana v [106]. Za učenje parametrov modela uporabimo Adam optimizator s stopnjo učenja 0,001. Učne podatke sproti, med samim učenjem (angl. *on the fly*) obogatimo (angl. *augment*), da bi preprečili pretirano prilagajanje podatkov. Model po 70 dobah (angl. *epoch*) doseže stabilno vrednost izgube. Ko je model enkrat naučen, na vhod sprejme 400×400 slikovnih elementov veliko sliko in vrne 512-dimenzionalno predstavitev na izhodu mreže (ki smo ji odstranili *softmax* sloja).

5.3.3. Ovrednotenje modela za segmentacijo beločnice

Naše eksperimente začnemo z ovrednotenjem prve faze našega razpoznavalnega cevovoda, ki segmentira regijo beločnice. Cilj te serije poizkusov je prikazati, kako naučena SegNet arhitektura deluje na problemu segmentacije beločnice in kako se primerja s konkurenčnimi segmentacijskimi modeli. Vredno je omeniti, da se napake, do katerih pride v tej fazi cevovoda, prenašajo skozi ves nadaljnji cevodov. Dokler je segmentiran večinski del beločnice, manjše napake (npr. malenkost zgrešena napoved meje med beločnico in

kožo, zaznavanje trepalnic v predelu beločnice itd.) niti niso usodne, saj jih lahko delno popravi tudi naslednja segmentacijska faza.

Da bi predstavili referenčni okvir uspešnosti arhitekture SegNet, smo implementirali 4 dodatne segmentacijske tehnike in jih preizkusili na naši testni množici. Bolj točno smo preizkusili delovanje treh vrhunskih (angl. state-of-the art) arhitektur za reševanje segmentacijskih problemov, ki temeljijo na konvolucijskih nevronskih mrežah ter en segmentacijski model, ki je bil izdelan posebej za segmentacijo beločnice. Bolj podrobne zgradbe globokih arhitektur so bile predstavljene v podpoglavju 3.2.4 – na tem mestu jih le na kratko povzamemo ter dodamo opis četrte segmentacijske metode:

- **RefineNet-50 in RefineNet-101:** RefineNet [95] je globoki segmentacijski model, ki je zgrajen okoli koncepta učenja s preostanki [109]. Glavna ideja je izrabiti značilke z vseh nivojev, s čimer dosežemo visoko ločljive semantične mape z grobofinim načinom (angl. course-to-fine manner). Odvisno od problema lahko naučimo različne globine mreže RefineNet. V našem delu smo uporabili dve različni možnosti, in sicer eno globine 50 (poimenujemo jo RefineNet-50) in drugo globine 101 (RefineNet-101). Modela učimo z istimi podatki in protokolom kot SegNet. Za eksperimente uporabljamo javno dostopno implementacijo ². Potrebno je omeniti, da smo s to mrežo leta 2018 dosegli zmagovalni rezultat na tekmovanju segmentacije beločnice, ki je bilo organizirano v sklopu mednarodne biometrične konference (angl. International Conference on Biometrics, ICB) [44].
- **UNet:** U-Net [96] je popularna segmentacijska arhitektura. Podobno kot arhitektura SegNet je sestavljena iz kodirnika in dekodirnika, s to razliko, da pretok informacije iz kodirnika v dekodirnik zagotavlja s konkatencijo map značilk iz kodirnika v istoležne dele dekodirnika. Model učimo na istih podatkih in po istem protokolu, kakor smo učili SegNet. Za naše eksperimente uporabljamo Keras implementacijo (z uporabo TensorFlow zaledja).
- **Unsupervised Sclera Segmentation (USS)** [110]: Za razliko od modelov, ki smo jih omenili zgoraj, USS predstavlja nenadzorovano (angl. unsupervised) segmentacijsko tehniko, ki ne temelji na uporabi predznanja. Tehnika operira na sivinskih slikah in je osnovana na adaptivni normalizaciji histograma, ki ji sledi gručenje ter adaptivno upragovanje. Več podrobnosti je možno najti v delu [110]. Tehnika je na tekmovanju SSERBC 2017 [20] osvojila drugo mesto. Da bi zagotovili pošteno primerjavo uspešnosti segmentacije, smo uporabili kodo, ki so nam jo priskrbeli avtorji.

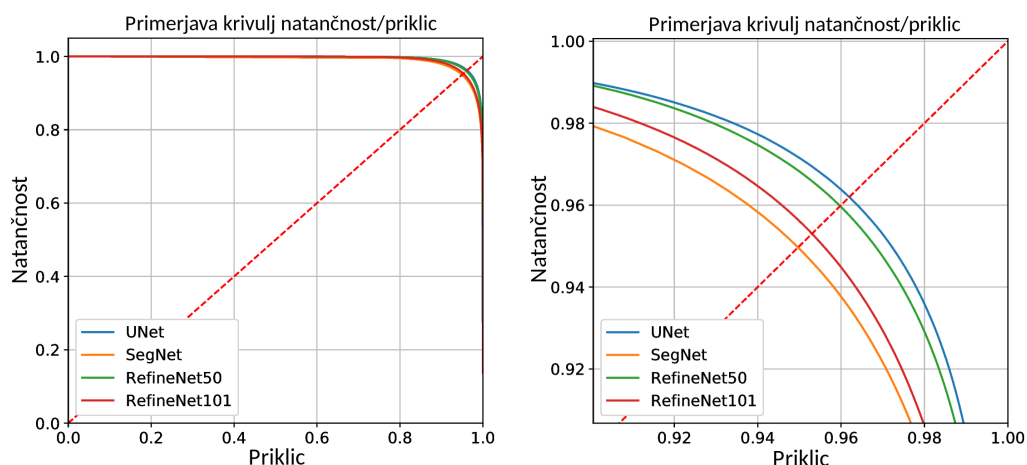
²Dostopno na: <https://github.com/guosheng/refinenet>

Tabela 5.2: Segmentacijski rezultati so pridobljeni na podlagi binarnih segmentacijskih mask. Za ovrednotenje CNN modelov so binarne maske pridobljene z upragovanjem dobljenih verjetnostnih map glede na prag $\Delta = 0,5$, medtem ko USS na izhod že vrne binarno masko segmentirane beločnice.

Algoritem	Natančnost	Priklic	F ₁ -ocena
UNet [96] (naš)	0,960	0,955	0,957
RefineNet-50 [95] (naš)	0,946	0,964	0,954
RefineNet-101 [95] (naš)	0,928	0,963	0,945
SegNet [20, 97] (naš)	0,858	0,986	0,916
Unsupervised Sclera Segmentation (USS) [20, 110]	0,735	0,711	0,712

Pomembno je opaziti, da trije CNN modeli kot končni rezultat vrnejo mape verjetnosti, medtem ko metoda USS kot končni rezultat vrne binarne maske. Da primerjavo med CNN modeli in USS lahko izvedemo, moramo verjetnostne mape CNN modelov binarizirati, kar naredimo z upragovanjem pri izbranem pragu $\Delta = 0,5$. V tabeli 5.2 beležimo natančnost, priklic in F₁-oceno za vse testirane metode. Ker vsi CNN modeli na svoj izhod vrnejo verjetnostne mape, jih lahko dodatno primerjamo z izrisom celotnih krivulj natančnost-priklic, ki so prikazane na sliki 5.5.

Rezultati pokažejo, da CNN modeli delujejo podobno dobro in da nobeden izmed modelov nima izrazitih prednosti pred drugimi. Po drugi strani pa USS pristop deluje nekoliko slabše, kar je konsistentno z rezultati, ki so bili predstavljeni na lestvici tekmovanja [20]. Zanimiva je opazka, da segmentacijski prag $\Delta = 0,5$ zagotavlja relativno podobno natančnost in priklic pri UNet in obeh RefineNet modelih, medtem ko je pri SegNet-u priklic precej večji od natančnosti (kar posledično pomeni nižjo F₁-oceno). Povprečen čas za segmentacijo ene slike (izračunano na množici 100 slik) je 1,2 sekunde za UNet, 0,6 sekunde za RefineNet-50, 0,8 sekunde za RefineNet-101, 0,15 sekunde za SegNet ter 0,34 sekunde za USS. V naši eksperimentih se je kot najhitrejši model med testiranimi modeli izkazal SegNet. Na sliki 5.6 prikažemo nekaj kvalitativnih rezultatov, ki smo jih dobili na množici testnih slik. Prvi stolpec prikazuje originalno RGB sliko očesa, drugi stolpec ročno anotacijo beločnice, preostali stolpci pa rezultate, ki smo jih pridobili z uporabljenimi modeli. Rezultati modelov si z leve proti desni sledijo v naslednjem zaporedju: USS, RefineNet-50, RefineNet-101, SegNet in UNet. Ti rezultati ponovno prikažejo, da vsi modeli, osnovani na CNN, zagotavljajo podobno uspešnost. V praksi bi bil za cevovod razpoznavanja primeren katerikoli izmed testiranih CNN segmentacijskih modelov. Za naš cevovod smo zaradi najhitrejšega izračunavanja izbrali SegNet.

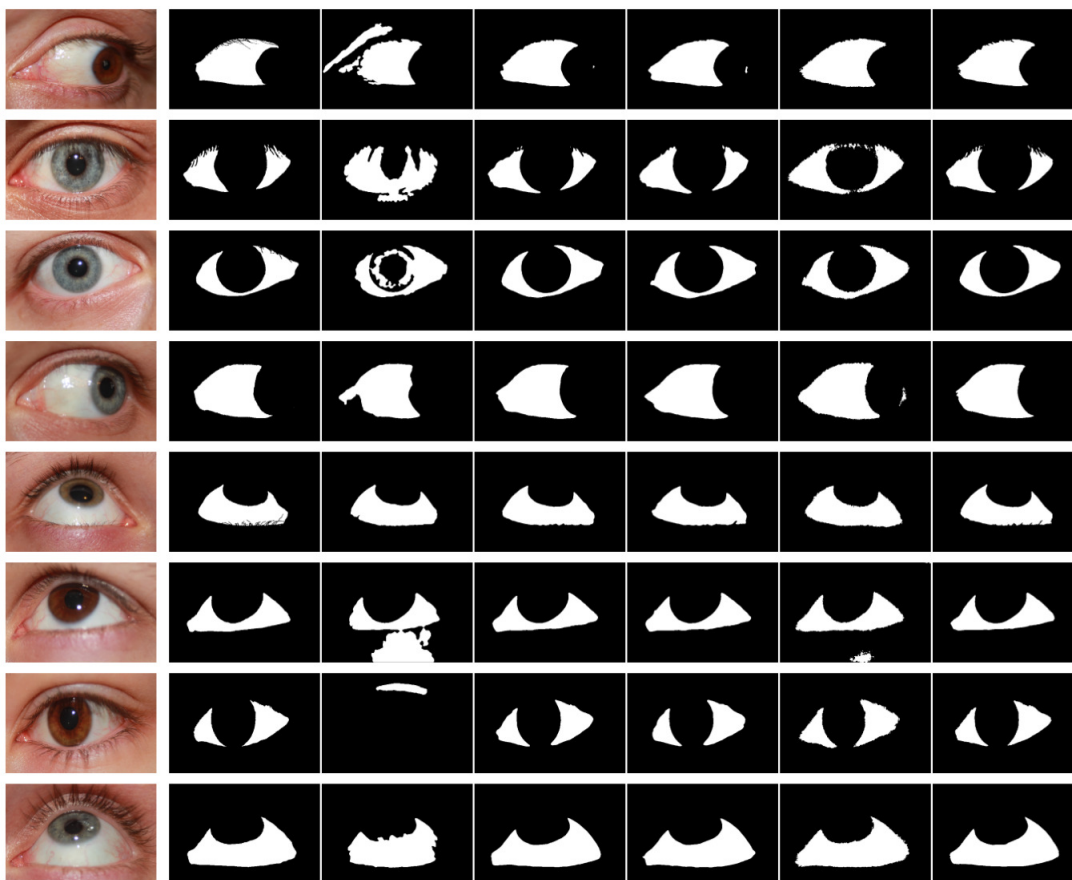


Slika 5.5: Prikaz krivulj natančnost-priklic za testirane CNN modele. USS tukaj ni prikazana, ker metoda vrne le binarno masko segmentirane beločnice. Levi graf kaže celoten izris krivulje, ki ga dobimo s premikanjem segmentacijskega praga Δ čez vse možne vrednosti, medtem ko desni graf prikazuje približno regijo levega grafa, kjer se vidijo majhne razlike med tehnikami. Rdeča pikčasta črta prikazuje operativne točke, kjer sta natančnost in priklic enaka.

5.3.4. Ovrednotenje modelov za segmentacijo vaskulature

V naslednjem sklopu eksperimentov ovrednotimo uspešnost druge segmentacijske faze našega cevovoda, ki je namenjena segmentaciji vaskularnih struktur iz beločnice. Vhod v to fazo je ponovno RGB slika očesa, le da je ta maskirana s segmentacijskim izhodom, ki ga vrne SegNet model iz prejšnjega podglavja. Opisano maskiranje je prikazano tudi na sliki 5.7. Poudarili smo že, da je druga segmentacijska faza narejena po principu razstavljanja osnovne slike na podslike ter na sestavljanju podrešitev v celotno rešitev (angl. patch-wise approach). S tem se izognemo izgubljanju informacije o manjših in tanjših žilah, do katerega bi prišlo, če bi v model za segmentacijo vaskularnih struktur dali le celo, nerazstavljeno sliko.

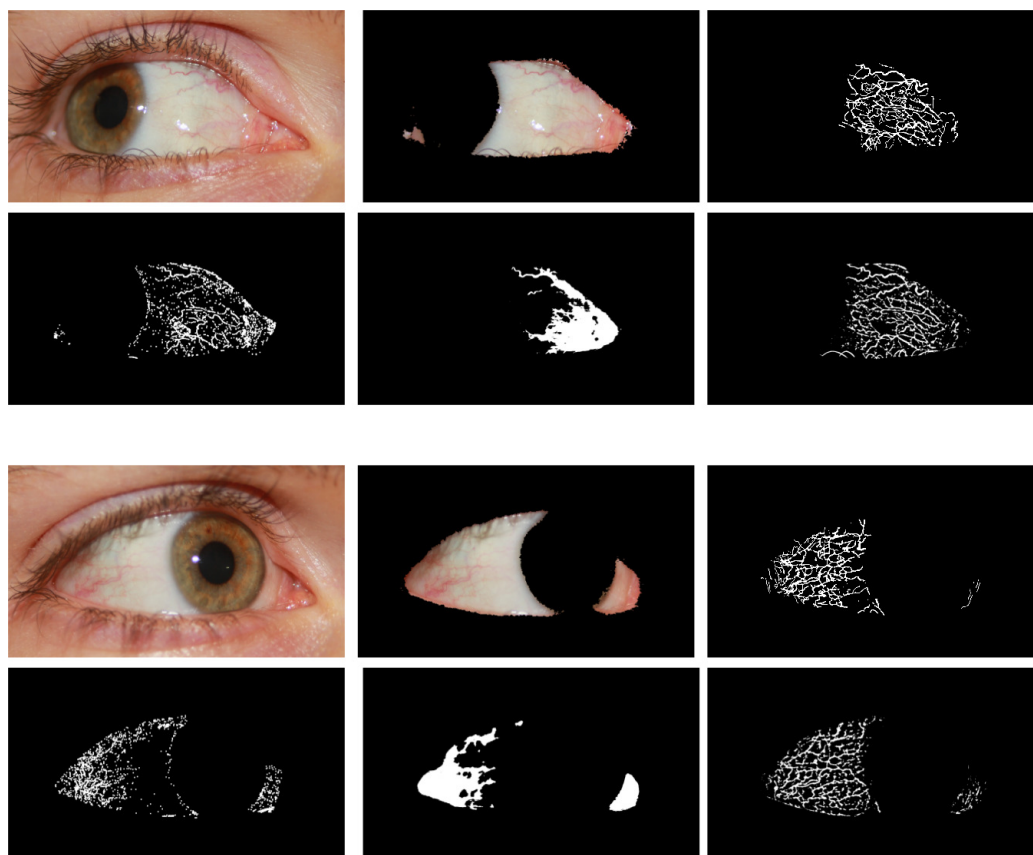
Predlagan SegNet model, ki ga v tem delu cevovoda uporabljamo za segmentacijo vaskulature, bomo primerjali z UNet modelom in z metodo Adaptivnega Gaussovega upravljanja [111]. Postopek učenja UNet modela je enak kakor pri SegNet-u, prav tako pa gre pri UNet modelu za enako strategijo razstavljanja vhodne slike na podslike in sestavljanje podrešitev v končno rešitev (verjetnostno mapo). Pri Adaptivnem Gaussovem upravljanju



Slika 5.6: Kvalitativna primerjava rezultatov segmentacije, ki jih vrnejo testirani segmentacijski modeli. Prvi stolpec prikazuje vhodno RGB sliko očesa, drugi ročno anotacijo beločnice, preostali stolpci pa segmentacijske rezultate pridobljene z (od leve proti desni): USS, RefineNet-50, RefineNet-101, SegNet in UNet. Opazimo, da CNN modeli (zadnji 4 stolpci) vrnejo zelo podoben segmentacijski rezultat in se razlikujejo le v manjših podrobnostih.

vanju upravljanje izvajamo brez strategije razstavljanja in sestavljanja neposredno nad celotno vhodno RGB sliko.

Ovrednotenje bomo izvedli tako, da bomo celotne rešitve CNN modelov SegNet in UNet (dobljene mape verjetnosti) binarizirali z upravljanjem. Pri navadnem Adaptivnem Gaussovem upravljanju je rezultat že privzeto binaren. Binarne napovedi vseh treh metod bodo nato primerjane z binarnimi ročnimi anotacijami, pri čemer bomo beležili natančnost, priklic in F_1 -oceno. Za binarizacijo verjetnostnih map SegNet modela bomo uporabljali Adaptivno Gaussovo upravljanje, za upravljanje verjetnostnih map UNet modela pa trivialno binarno upravljanje. Dva različna postopka za upravljanje CNN



Slika 5.7: Dva primera rezultatov segmentacije vaskularnih struktur. Šest podslik pri posameznem primeru prikazuje (z leve proti desni, od zgoraj navzdol): vhodno RGB sliko, maskirana beločnica (ki jo dobimo kot izhod prve segmentacijske faze v cevovovu), ročna anotacija vaskularnih struktur, rezultati adaptivnega Gaussovega upragovanja, rezultati modela UNet ter rezultati modela SegNet.

modelov izberemo zato, ker smo v naših predhodnih eksperimentih za kombinacijo UNet + Adaptivno Gaussovo upragovanje dobili slabše kvalitativne rezultate kakor z uporabo kombinacije UNet + Binarno upragovanje. To pripisujemo dejstvu, da UNet model vrača relativno nejasno mejo med žilami in okoljem. RefineNet modelov v eksperiment ne vključujemo, saj nam predhodni rezultati nakazujejo na to, da nimamo dovolj podatkov, s katerimi bi lahko naučili dovolj kompetitiven model (kar je verjetno posledica globine RefineNet mrež). Rezultati eksperimentov so predstavljeni v tabeli 5.3. Kakor je možno videti, SegNet z velikim presežkom zagotavlja najboljše rezultate, pri čemer je F_1 -ocena 0,649. UNet doseže F_1 -oceno 0,351, metoda adaptivnega upragovanja pa 0,316, kar sta precej slabša rezultata od F_1 -ocene SegNet-a.

Tabela 5.3: Primerjava tehnik za segmentacijo vaskularnih struktur. Predstavljeni so rezultati, ki jih dosežeta modela SegNet in UNet ter metoda adaptivnega upragovanja. Opazimo, da SegNet z veliko prednostjo doseže najboljši rezultat.

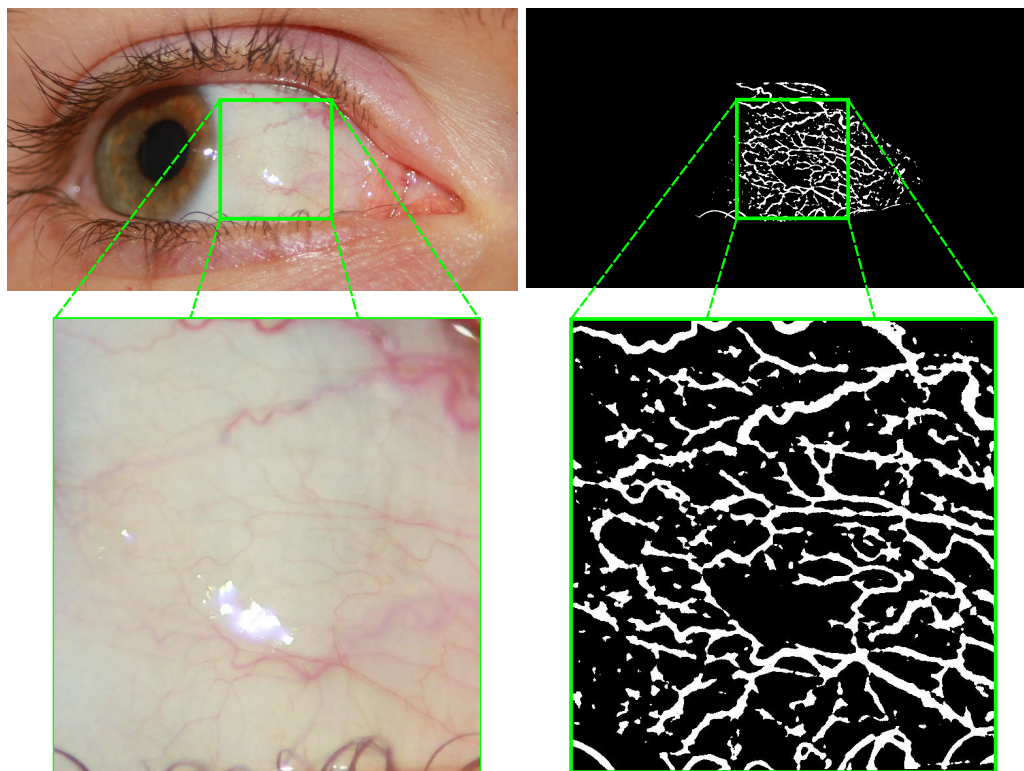
Algoritem	Natančnost	Priklic	F ₁ -ocena
SegNet [31] + Adaptivno Gaussovo upragovanje (naš)	0,678	0,622	0,649
UNet [96] + Binarno upragovanje (naš)	0,304	0,459	0,351
Adaptivno Gaussovo upragovanje	0,322	0,355	0,316

Kvalitativni segmentacijski rezultati so v vizualni obliki predstavljeni na sliki 5.7. Prikazana sta dva primera, pri čemer vsak primer sestavlja 6 podslik, ki pomenijo naslednje (od leve proti desni, od zgoraj navzdol): vhodna slika, maskirana regija beločnice, ročna anotacija vaskularnih struktur, rezultat adaptivnega Gaussovega upragovanja, rezultat modela UNet ter rezultat modela SegNet. Zanimivo je videti, kakšen nivo podrobnosti je zmožen zaznati SegNet model. Kljub slabemu kontrastu nekaterih bolj finih žil, jih model večino še vedno uspešno segmentira. Z UNet-om v naših eksperimentih nismo uspeli doseči kvalitetnih rezultatov, saj model v glavnem pravilno zaznava le večje žile, hkrati pa k žilam napačno pripiše tudi veliko njihovih sosednjih slikovnih elementov.

Da bi bolje prikazali, kakšno segmentacijsko kvaliteto dobimo z uporabo SegNet modela, je na sliki 5.8 prikazana približna slika vaskularne strukture očesa s korespondenčnim segmentacijskim rezultatom. Vidimo, da model uspešno segmentira večino vaskularne strukture. Kot žile pa napačno zazna tudi nekatere trepalnice, katerih vzorci so v veliki meri podobni vaskularnim strukturam.

Ker je segmentacija izvedena po principu razstavljanja in sestavljanja podrešitev v celotno rešitev, je povprečen čas, ki ga potrebujemo za procesiranje ene slike z uporabo ene grafične kartice, 5,58 sekund. Vredno je poudariti, da se z uporabo večjega števila grafičnih kartic ta računski korak lahko paralelizira, saj se posamezne podslike lahko procesira neodvisno, kar lahko čas procesiranja občutno zmanjša. Celotna pridobitev regije interesa (segmentacija beločnice skupaj s segmentacijo vaskulature) traja torej približno 6 sekund, če uporabljamo le eno grafično kartico.

V prejšnjem podpoglavju smo videli, da z arhitekturo UNet lahko kvalitetno segmentiramo regijo beločnice, pri čemer je bil segmentacijski rezultat mreže UNet primerljiv rezultatu, ki ga na izhod vrne SegNet. V tem podpoglavju je uspešnost UNet modela (glede na prejšnje podpoglavje) občutno slabša, kar pripisujemo trem dejavnikom: *i*) kontrast žil glede na beločnico je manjši kot kontrast beločnice glede na ostale dele očesa, *ii*)

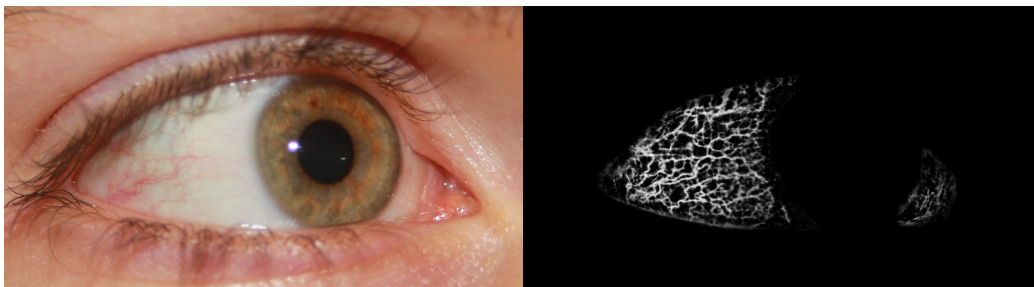


Slika 5.8: Vizualizacija drobnih vaskularnih struktur, ki jih segmentacijski model najde na sliki. Slika prikazuje približno regijo vaskularne strukture v očesu (na levi) ter korepondenčni binarizirani izhod našega modela (na desni.)

oblika žil je bolj kompleksna kot oblika beločnice in *iii*) globina arhitekture UNet morda ni dovolj globoka za uspešno segmentacijo žil. Pri segmentaciji beločnice smo morali iz ene velike slike segmentirati beločnico, ki ima precej kontrastno mejo z ostalimi deli očesa (npr. rob beločnice s šarenico ali vekami). V tem podpoglavju moramo najti žile na večjem številu podslik, pri čemer so na posamezni podsliki vidni le manjši odseki žil, ki jih je mnogokrat težko opaziti, zaradi pogosto šibkega kontrasta med žilami in beločnico. Beločnične regije (ki so bile segmentacijski objekt v prejšnjem podpoglavju) imajo tipično večjo segmentacijsko površino kot žile, ki so pogosto tanke in kompleksno razvejane. Razlog za slabše delovanje UNet na problemu segmentacije žil bi bila lahko tudi nedovoljšnja globina arhitekture (ki morda zadostuje za manj kompleksen problem segmentacije beločnice). Naši rezultati namreč kažejo na to, da s SegNet (ki je precej globlja mreža) lahko dobimo precej boljše rezultate.

Ti rezultati kažejo na to, da lahko z naučenim SegNet modelom pridobimo dovolj kvalitetne segmentacijske rezultate, ki so nadalje uporabni za namene razpoznavanja. V

naslednjem podpoglavju na podlagi pridobljenih segmentacij vaskularnih struktur ovrednotimo uspešnost delovanja našega razpoznavalnega pristopa.



Slika 5.9: Primer vhodne slike s korespondenčno verjetnostno mapo žil, ki jo vrne SegNet.

5.3.5. Eksperimenti razpoznavanja

V zadnji vrsti eksperimentov v tem poglavju smo ovrednotili uspešnost celotnega segmentacijskega cevovoda, rezultate segmentiranih vaskularnih struktur pa smo dali na vhod modela ScleraNET, ki je iz njih izluščil diskriminativne značilke. Ponovno opozarjamo, da so na vhod ScleraNET podane verjetnostne mape žil in ne binarna maska vaskulature. Verjetnostna mapa je sivinska slika, kjer svetlejša vrednost slikovnega elementa pomeni višjo verjetnost vaskularne strukture. Primer verjetnostne mape, ki jo pridobimo s SegNet modelom, je prikazana na sliki 5.9.

Ko so diskriminativne značilke vhodne slike poračunane, je za izračun ocene ujemanja uporabljena kosinusna podobnost (angl. cosine similarity), ki jo opisuje enačba (5.9). Sklepamo torej, da beločnica na vhodni sliki pripada tisti identiteti, s katero ima najvišje ujemanje.

$$\text{podobnost}(x, y) = \frac{xy}{|x||y|} = \frac{\sum_{i=0}^{n-1} x_i y_i}{\sqrt{\sum_{i=0}^{n-1} x_i^2} \times \sqrt{\sum_{i=0}^{n-1} y_i^2}} \quad (5.9)$$

Z uporabo protokola vsi-proti-vsem (angl. all-vs-all) smo za ovrednotenje ScleraNET naredili dva tipa razpoznavalnih eksperimentov in sicer:

- *Verifikacijski eksperimenti:* Pri tem tipu eksperimentov je za vsako sliko iz testne množice poračunano ujemanje z vsemi drugimi slikami v testni množici. Napako legitimnih verifikacijskih poizkusov merimo s primerjavo podobnosti ujemajočih se razpoznavalnih identitet, napako nelegitimnih verifikacijskih poskusov pa s primerjavo podobnosti različnih razpoznavalnih identitet. Glede na izbrano testno množico SBVPI naredimo torej vse skupaj 29.340 legitimnih in 2.627.560 nelegitimnih

verifikacijskih poskusov. Poudariti je potrebno, da gre za zelo zahteven eksperimentalni scenarij, saj sta na posameznih slikah v paru slik (med katerimi računamo ujemanje) pogosto prikazani dve drugačni smeri pogledov. Očesi, ki npr. pripadata isti identiteti, pri čemer eno gleda levo in drugo desno, skoraj nimata vidnih delov vaskulature, ki bi se prekrivali.

- *Identifikacijski eksperimenti:* Pri tem tipu eksperimentov je vsaka slika iz testne množice enkrat mišljena kot poskusna slika (angl. probe) in se poročuna ujemanje z vsemi ostalimi slikami v testni množici. Po našem protokolu imamo torej 1.630 identifikacijskih poizkusov, pri katerih moramo izračunati podobnost z ostalimi 1.629 slikami iz testne množice. V galeriji imamo za isto identiteto zbranih več slik in obenem je za isto identiteto zbranih več različnih smeri pogleda. Zato se za izračun ujemanja med dvema identitetama dejansko izračuna več ocen ujemanja (med vsemi možnimi pari slik). Za vsako poskusno sliko se končna napoved identitete določi tako, da se pogleda, s katero od preostalih slik v testni množici ima najvišje ujemanje. Gledano z zornega kota implementacije je vseh možnih učnih identitet v testni množici 100, pri čemer po dve tovrstni identiteti kažeta na isto razpoznavalno identiteto (levo in desno oko iste osebe). Razpoznavalnih identitet je 50.

Ker slike očeš niso poravnane, kot izhodišče implementiramo razpoznavalni algoritem, ki je osnovan na metodi SIFT [112], kot vhod v algoritem pa ponovno uporabimo verjetnostne mape beločnične vaskulature. Za vsak slikovni par za izračun ujemanja vzamemo povprečno Evklidsko razdaljo med ujemajočima se SIFT deskriptorjema. Ker je pristop s SIFT lokalni ter osnovan na korespondenčnih ključnih točkah, je še posebej primeren za reševanje tega problema, saj morajo biti (delno vidne) neporavnane strukture različnih pogledov primerjane med seboj.

Iz rezultatov, ki so predstavljeni v tabeli 5.4 ter na slikah 5.10 in 5.11, pri čemer so rezultati za ScleraNET označeni kot CNN, lahko vidimo, da s predlaganim pristopom dobimo AUC 0,943 za verifikacijski eksperiment, v primerjavi s SIFT, ki doseže AUC 0,837. Z EER, ki je enak 0,108, ScleraNET zagotovi mnogo manjši ERR kakor SIFT. Rezultati CNN so vzpodbudni, saj verifikacijski eksperimenti vključujejo najbolj zahtevne verifikacijske poskuse med različnimi vrstami smeri pogledov.

Za identifikacijske eksperimente rezultati ScleraNET v razpoznavanju po protokolu Rank-1 dosežejo razpoznavalno oceno 0,727 ter Rank-5 oceno 0,909, v primerjavi s SIFT, ki doseže Rank-1 enak 0,563 in Rank-5 enak 0,681. Razlika v uspešnosti med metodami je velika za vse operativne točke na CMC krivuljah in ponovno se izkaže, da CNN pristop deluje bolje kakor SIFT. Razlog za takšno razliko pripisujemo sposobnosti CNN, ki se nauči, katera smer pogleda je prikazana na slikah. Za razliko od SIFT zaradi tega bolje

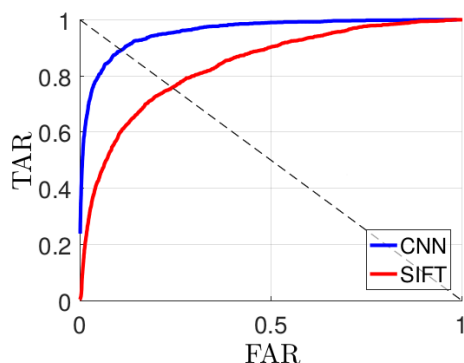
izlušči diskriminativno informacijo, ki jo vsebujejo vaskularne strukture.

Tabela 5.4: Rezultati eksperimentov razpoznavanja na podlagi vaskularnih struktur. Rezultati so ločeno prikazani za verifikacijske in identifikacijske eksperimente. Predlagani ScleraNET model zagotavlja kompetitivno uspešnost in je znatno boljši od SIFT algoritma, ki smo ga vzeli kot izhodišče.

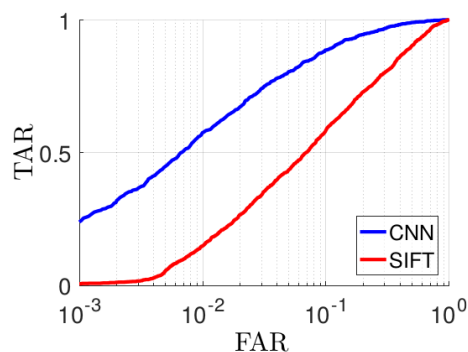
Ekspiriment	Algoritem	VER@0.1FAR	VER@1FAR	EER	AUC
Verifikacija	ScleraNET (naš)	0,239	0,575	0,108	0,943
	SIFT	0,006	0,149	0,242	0,837
	Algoritem		Rank-1	Rank-5	AUCMC
Identifikacija	ScleraNET (naš)		0,727	0,909	0,981
	SIFT		0,563	0,681	0,892

5.4. Diskusija

V tem poglavju smo predlagali celoten cevovod za razpoznavanje na podlagi beločnice, ki temelji na metodah globokega učenja. Pri tem smo ovrednotili posamezne faze tega cevovoda v primerjavi z obstoječimi, konkurenčnimi algoritmi. Naš postopek za segmentacijo beločnice z globokim učenjem zagotavlja vrhunske rezultate, ki po uspešnosti precej odstopajo od metod, ki so jih predlagali drugi avtorji. Pri naših rezultatih se izkaže, da beločnico lahko uspešno segmentiramo z uporabo različnih arhitektur celostno-konvolucijskih arhitektur. V praksi lahko med ovrednotenimi CNN arhitekturami izberemo poljubno arhitekturo, saj le-te zagotavljajo zelo podobne segmentacijske uspešnosti, ki se razlikujejo le v manjših podrobnostih. Druga faza za pridobitev regije interesa je segmentacija žil iz beločnice, kjer smo najboljše rezultate dosegli z uporabo SegNet modela. Primerjavo smo naredili tudi z arhitekturo UNet, pri čemer se je izkazalo, da z arhitekturo UNet tokrat nismo uspeli doseči konkurenčnih rezultatov. Naslednja faza v našem cevovodu je luščenje značilke, kjer predlagamo svojo CNN arhitekturo, ki smo jo poimenovali ScleraNET. Naši eksperimenti pokažejo, da z uporabo ScleraNET lahko izluščimo bolj diskriminativne značilke, kot jih omogoča uporaba tradicionalnega algoritma SIFT. V naših predhodnih eksperimentih se je izkazalo, da je boljše, če na vhod ScleraNET namesto binarnih napovedi podamo verjetnostne mape žil. Če povzamemo, smo razvili cevovod, ki ima potencial izboljšati natančnost sistemov, ki razpoznavajo na podlagi drugih očesnih modalnosti.

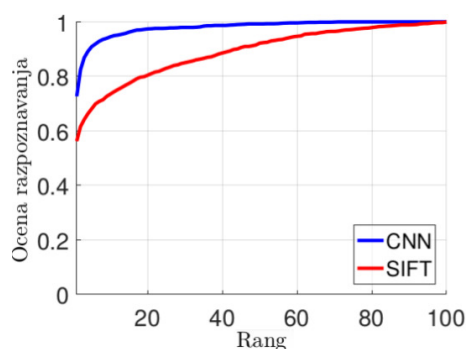


(a) Na linearni skali prikazani ROC krivulji.

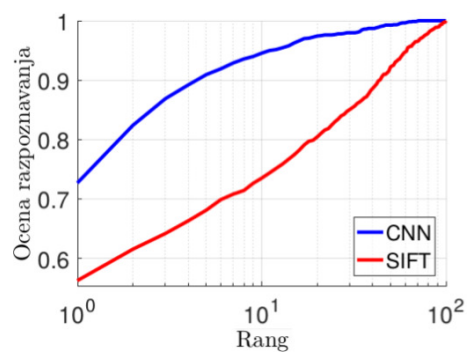


(b) Na semi-logaritmični skali prikazani ROC krivulji.

Slika 5.10: Rezultati verifikacijskih eksperimentov razpoznavanja iz beločnice. Na levem grafu sta ROC krivulji prikazana na linearni skali, medtem ko desni graf prikazuje ROC krivulji na semi-logaritmični skali, zaradi česar je bolj vidna razlika pri nižjih FAR vrednostih. Iz primerjave lahko vidimo, da ScleraNET (označena kot CNN) z veliko razliko deluje boljše, kakor pristop na podlagi SIFT.



(a) Na linearni skali prikazani CMC krivulji.



(b) Na semi-logaritmični skali prikazani CMC krivulji.

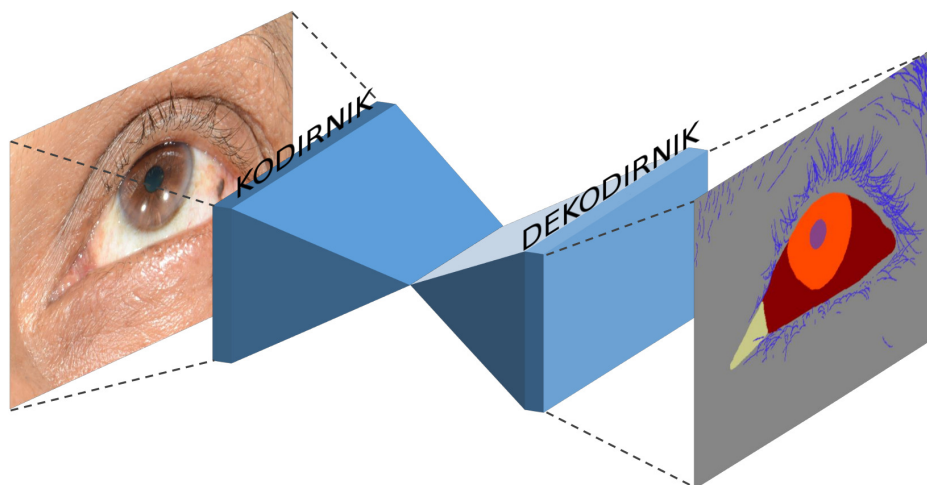
Slika 5.11: Rezultati identifikacijskih eksperimentov razpoznavanja iz beločnice. Na levem grafu sta CMC krivulji prikazani na linearni skali, medtem ko desni graf prikazuje CMC krivulji na semi-logaritmični skali, zaradi česar je bolj vidna razlika pri nižjih rangih. Iz primerjave lahko vidimo, da ScleraNET (označena kot CNN) z veliko razliko deluje boljše, kakor pristop na podlagi SIFT.

Poglavje 6

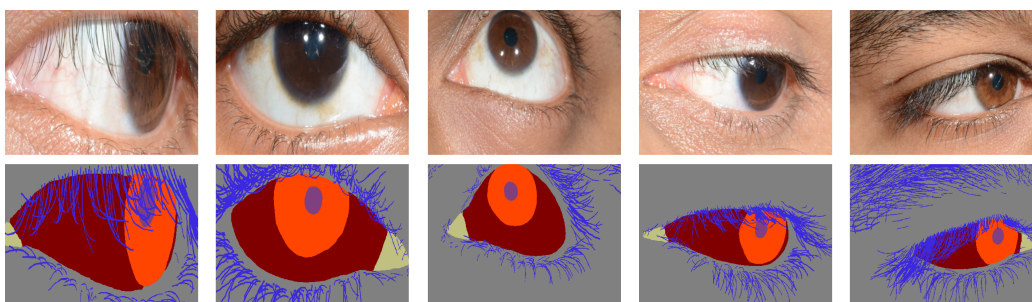
Večrazredna segmentacija očesa

6.1. Uvod

V prejšnjem poglavju smo videli, da potrebujemo za pridobitev značilk, na podlagi katerih bomo zmožni kvalitetno izračunati ujemanje med dvema beločnicama, dobre segmentacijske modele. To velja tudi za cevovode, ki razpoznavanje na podlagi drugih očesnih modalnosti. Za cevovod, ki razpozna izključno na podlagi šarenice, moramo naučiti model, ki segmentira šarenico. Na enak način potrebujemo drug segmentacijski model za cevovod, ki razpozna na podlagi periokularnega dela. Ker se v našem delu ukvarjamo s fuzijo treh očesnih modalnosti, se pojavi vprašanje, ali lahko namesto treh modelov, kjer vsak segmentira svojo modalnost, naučimo le en model, ki bi bil zmožen sliko očesa segmentirati v več razredov. Z uporabo večrazredne segmentacije bi se kompleksnost cevovoda zmanjšala, hkrati pa bi lahko zmanjšali računsko zahtevnost cevovoda. Dodatni razlog, da je večrazredna segmentacija pri intra-okularnem fuzijskem sistemu zaželjena, je dejstvo, da so vse modalnosti vidne na isti vhodni sliki (del slike je beločnica, del šarenica, del periokularni del itd.). V tem poglavju bomo izdelali segmentacijski algoritem, ki bo zmožen vhodno sliko očesa segmentirati v šest razredov. Vsebino tega poglavja smo objavili tudi v delu [97].



Slika 6.1: Uporabljamo segmentacijsko arhitekturo SegNet, ki jo sestavlja iz dveh konvolucijskih delov, in sicer kodirnika ter dekodirnika. Naš pristop je z enim samim segmentacijskim modelom iz slike zmožen segmentirati 6 razredov: zenico, šarenico, beločnico, trepalnice, medialni kantus ter periokularno regijo.



Slika 6.2: Primer slik oči (zgornja vrstica) in korespondenčnih anotacij na nivoju slikovnih elementov (spodnja vrstica), ki so uporabljene v tem delu. Barvno kodirane anotacije označujejo šest različnih razredov, in sicer: zenico, šarenico, beločnico, trepalnice, medialni kantus ter periokularno regijo. Uporabljene slike so vzete iz zbirke MASD [20].

6.2. Metodologija

V nadaljevanju je opisan naš pristop za večrazredno segmentacijo očesa. Kakor je ilustri-rano na sliki 6.1, za segmentacijo vhodnih slik oči uporabljamo arhitekturo konvolucij-skega kodirnika-dekodirnika. CED je zmožen slikovne elemente na sliki očesa klasificirati v

šest različnih razredov: zenico, šarenico, beločnico, trepalnice, medialni kantus in periokularno regijo. CED temelji na nedavno predstavljeni arhitekturi SegNet [31], ki je opisana v podpoglavju 3.2.4.

6.3. Večrazredna segmentacija in učenje

Da bi bili zmožni naučiti SegNet model za večrazredno segmentacijo, smo najprej ročno anotirali manjši del zbirke MASD (120 slik). Na nivoju slikovnih elementov (angl. pixel) smo tako, kot je prikazano na sliki 6.2, označili 6 razredov ($C = 6$), in sicer: zenico, šarenico, beločnico, trepalnice, medialni kantus ter periokularno regijo. Anotirane slike uporabimo za učenje parametrov SegNet modela, ki ga učimo z uporabo kategorične prečne entropije (angl. categorical cross-entropy). Ker so razredi na različnih slikah različno zastopani (na slikah večinoma prevladujejo slikovni elementi periokularnega dela, šarenice in beločnice), za uravnoteženje uporabljamo frekvenco mediane (angl. median frequency balancing), kakor je tudi predlagano v delu [31]. Ko je model enkrat naučen, mu na vhod podamo RGB sliko očesa $I(x, y) \in \mathbb{R}^{w \times h \times 3}$, ki jo sestavlja $w \times h$ slikovnih elementov. Izhod je šest verjetnostnih porazdelitev, pri čemer vsaka izmed njih pripada enemu izmed $C = 6$ razredov. To pomeni, da mreža za vsako lokacijo $s = [x, y]^T$ na vhodni sliki $I(x, y)$ na izhod vrne porazdelitev $\mathbf{p}_s = [p_{\omega_1}, \dots, p_{\omega_C}]^T \in \mathbb{R}^{C \times 1}$, kjer p_{ω_i} označuje verjetnost, da slikovni element na lokaciji s pripada i -temu ciljnemu razredu ter hkrati velja $\sum_{i=1}^C p_{\omega_i} = 1$. Za pridobitev končnega segmentacijskega rezultata lahko uporabimo eno izmed dveh strategij:

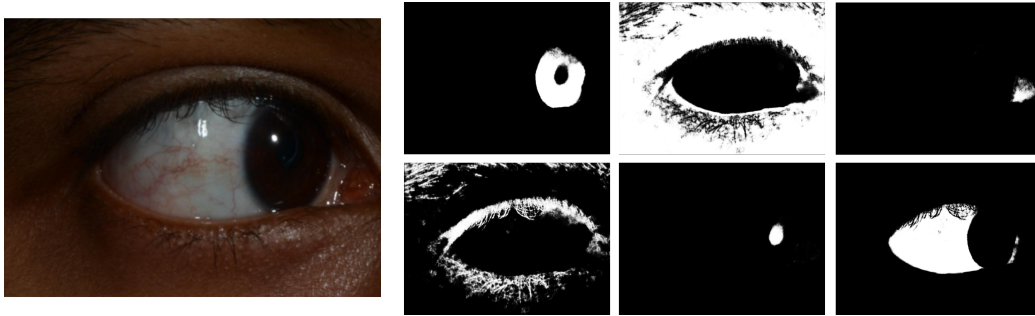
- *Strategija izbire maksimalnega elementa (angl. max-out strategy, MOS)*: Za vsako pozicijo slikovnega elementa s na vhodni sliki napovemo ciljni razred $\hat{\omega}_k \in \Omega$ tako, da pogledamo, katera izmed verjetnostnih map, ki jih je vrnil model, ima na lokaciji s najvišjo vrednost verjetnosti. Vsaka verjetnostna mapa namreč napoveduje določen razred. Matematično to lahko zapišemo kot:

$$\hat{\omega}_k = \arg \max_{\omega_i} (\mathbf{p}_s). \quad (6.1)$$

Ta strategija označi slikovni element na poziciji s ne glede na dejansko verjetnost (tudi če je verjetnost razreda majhna ali primerljiva z verjetnostjo drugega razreda). Na ta način se vsakemu slikovnemu elementu priredi natanko en ciljni razred.

- *Strategija upravanja (angl. thresholding strategy, THS)*: Vsako izmed šestih verjetnostnih map iz verjetnostne porazdelitve \mathbf{p}_s , ki jo vrne model, obravnavamo ločeno na sledeč način:

$$P_{\omega_i}(s) = p_{\omega_i}^{(s)}, \quad (6.2)$$



Slika 6.3: Primer verjetnostnih map $P_{\omega_i}(s)$, ki jih vrne *softmax* sloj modela SegNet. Slika prikazuje vhodno sliko (na levi strani) in korespondenčne verjetnostne mape (z leve proti desni, od zgoraj navzdol): šarenica, periokularna regija, medialni kantus, trepalnice in obrvi, zenica in beločnica.

pri čemer oznaka (s) označuje, da verjetnost p_{ω_i} pripada lokaciji slikovnega elementa $s = [x, y]^T$ in $x = 1, \dots, w$, $y = 1, \dots, h$. Primeri verjetnostnih map $P_{\omega_i}(s)$, ki so generirane za vhodno sliko, so prikazane na sliki 6.3. Končen segmentacijski rezultat (segmentacijska maska) je pridobljen z upragovanjem verjetnostnih map, pri čemer za vsak razred posebej določimo svojo vrednost praga Δ_{ω_i} , tako da velja $P_{\omega_i}(s) > \Delta_{\omega_i}$. Ker za vsak ciljni razred ločeno nastavimo vrednost praga, se lahko zgodi, da nekaterim slikovnim elementom ne določimo nobenega izmed ciljnih razredov, medtem ko se nekaterim slikovnim elementom določi več ciljnih razredov.

Ker nas zanima večrazredna segmentacija očesa, ročno anotiramo 120 slik na nivoju slikovnih elementov (ki pripadajo 30 različnim osebam), kakor je prikazano na sliki 6.2. To nam omogoča, da naučimo mrežo SegNet večrazredne segmentacije in ovrednotimo uspešnost le-te.

Pred učenjem SegNet mreže pomanjšamo vse učne slike na fiksno velikost 480×360 slikovnih elementov. Ker za ovrednotenje uporabljamo 4-kratno prečno preverjanje, vsakokrat za učenje uporabimo 90 slik ter 30 slik za testiranje. Ker 90 slik ne predstavlja dovolj velike količine podatkov, s katerimi bi lahko uspešno naučili model, ki bi bil sposoben dobro posploševati, ob vsakem izmed štirih učenj učno množico slik obogatimo (angl. augmentation). To naredimo po postopku, ki je predlagan v [101]. Iz vsake slike v učni množici z operacijami zrcaljenja, izrezovanja (angl. cropping), zamegljevanja, spremembami osvetlitve ter z uporabo afinih transformacij ustvarimo 100 novih slik. Naši eksperimenti kažejo, da s tovrstno obogatitvijo naših relativno majhnih učnih množic lahko naučimo odporne SegNet modele, ki dobro posplošujejo na še ne-videnih primerih.

Za učenje modela smo uporabili Caffe implementacijo mreže SegNet iz [31]. Model smo naučili na grafični kartici GTX 1080 Ti. Mrežo smo učili s 30.000 iteracijami, paketno

velikostjo 4 ter uĉnim korakom 0,001.

Tabela 6.1: Rezultati veĉrazredne segmentacije MOS strategije. Predstavljene so povpreĉne metrike uspešnosti in pripadajoĉe standardno odstopanje.

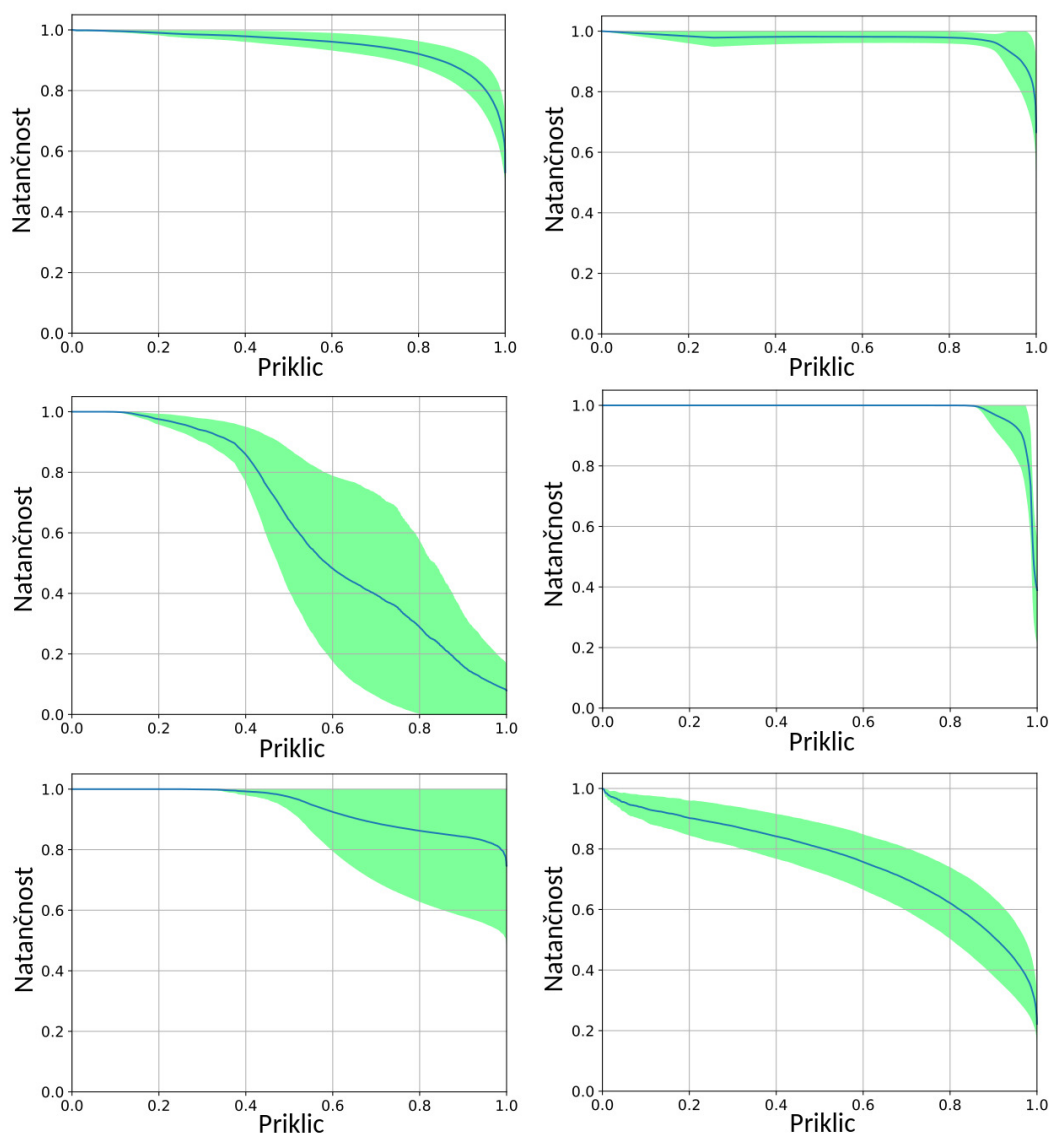
Razred (regija oĉesa)	Natanĉnost	Priklic	F ₁ -ocena
Šarenica	0,94 ± 0,05	0,89 ± 0,12	0,91 ± 0,10
Beloĉnica	0,92 ± 0,04	0,90 ± 0,09	0,91 ± 0,07
Zenica	0,89 ± 0,16	0,84 ± 0,21	0,85 ± 0,19
Periokularna regija	0,92 ± 0,05	0,89 ± 0,06	0,90 ± 0,05
Trepalnice	0,57 ± 0,10	0,72 ± 0,12	0,63 ± 0,10
Kantus	0,67 ± 0,28	0,45 ± 0,25	0,49 ± 0,24

6.4. Rezultati veĉrazredne segmentacije

V naših prvih eksperimentih ovrednotimo uspešnost modela SegNet za reševanje veĉrazredne segmentacije z uporabo 4-kratnega preĉnega preverjanja, po protokolu, ki smo ga natanĉneje opisali v prejšnjem podpoglavju. V tabeli 6.1 beležimo rezultate za MOS strategijo v obliki povpreĉne natanĉnosti, priklica in F₁-ocene s korespondenĉnimi standardnimi odstopanji, za THS strategijo pa v obliki krivulj natanĉnost-priklic, ki so prikazane na sliki 6.4. Opazimo, da pri THS strategiji lahko izrišemo celotno krivuljo natanĉnost-priklic, medtem ko pri MOS strategiji lahko beležimo le rezultate za posamezne toĉke.

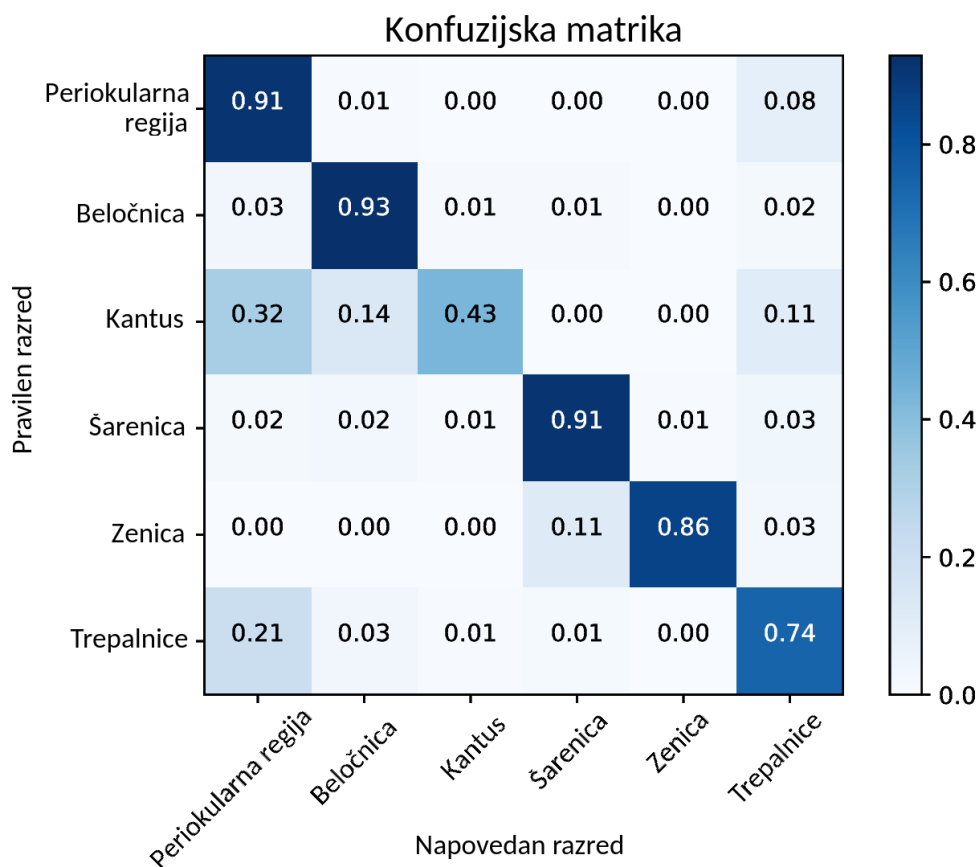
Rezultati prikazujejo, da so šarenica, beloĉnica in periokularna regija precej natanĉno segmentirane. Segmentacija zenice je malenkost slabša, vendar opazimo, da imamo pri meritvah oprava z visokim standardnim odstopanjem. Segmentacijska uspešnost trepalnic ter medialnega kantusa je občutno nižja.

Da bi dobili boljši vpogled v tipe napak, ki jih naredi SegNet model z MOS strategijo, na sliki 6.5 predstavimo normalizirano konfuzijsko matriko. Matrika je normalizirana za vsak del oĉesa posebej, kar pomeni, da je vsota vseh vrednosti v posamezni vrstici enaka 1. Matrika prikazuje, katere razrede model najbolj pogosto pomeša med seboj. Vidimo, da je periokularna regija najveĉkrat napaĉno klasificirana kot trepalnice, kantus je zamešan za periokularni del in v manjšem obsegu za beloĉnico, trepalnice so pogosto napaĉno klasificirane kot periokularna regija in zenica je napaĉno klasificirana kot šarenica. Kljub



Slika 6.4: Povprečne krivulje natančnost-priklíc (s standardnim odstopanjem), ki jih izrišemo med ovrednotenjem SegNet modela z uporabo THS strategije in 4-kratnega prečnega preverjanja. Grafi prikazujejo rezultate (od leve proti desni, od zgoraj navzdol): periokularne regije, beločnice, kantusa, šarenice, zenice ter trepalnic.

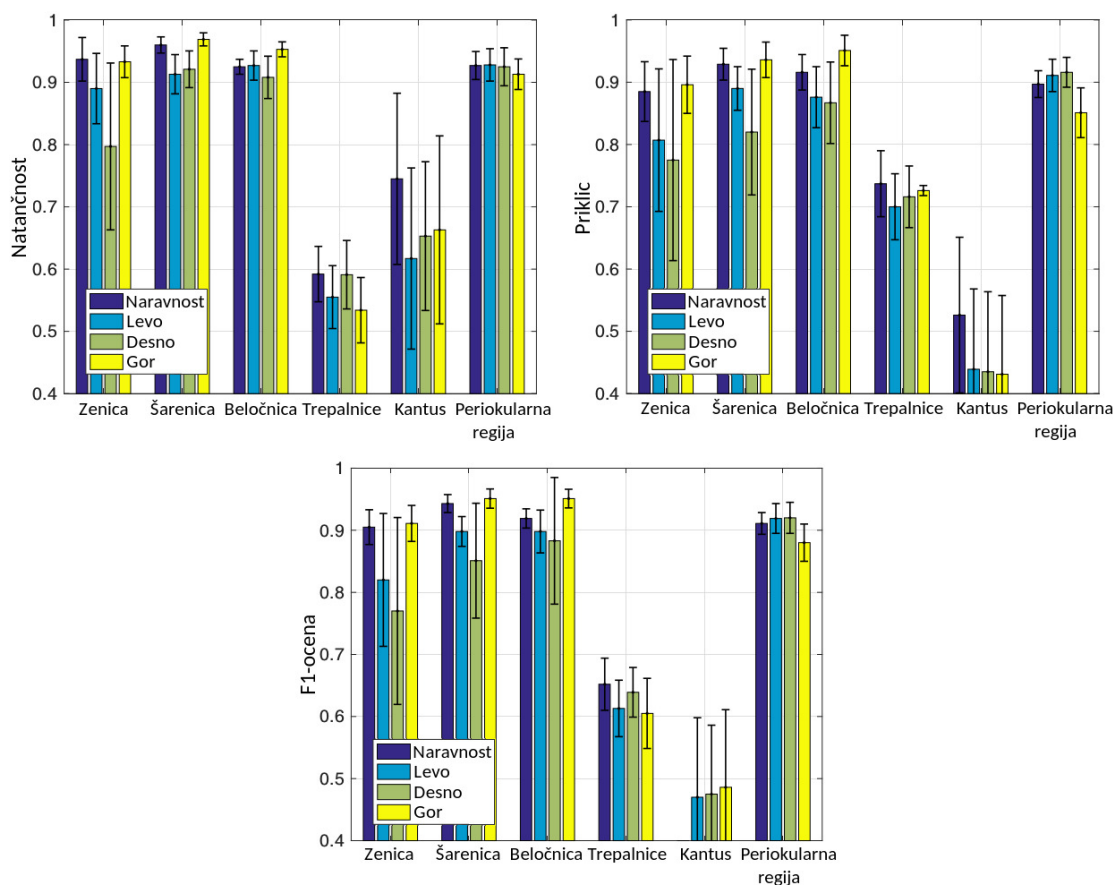
omenjenim napakam je na prikazani konfuzijski matriki jasno poudarjena diagonala (konfuzijska matrika popolne klasifikacije bi imela na diagonali verjetnosti 1, ostale verjetnosti pa 0), kar pomeni, da so napake za večino razredov relativno majhne. Regiji beločnice in šarenice imata napake glede na ostale razrede približno enako razporejene.



Slika 6.5: Konfuzijska matrika MOS strategije. Matrika prikazuje tipe napak, ki jih ob napovedovanju naredi SegNet model.

6.5. Analiza občutljivosti in kvalitativni rezultati

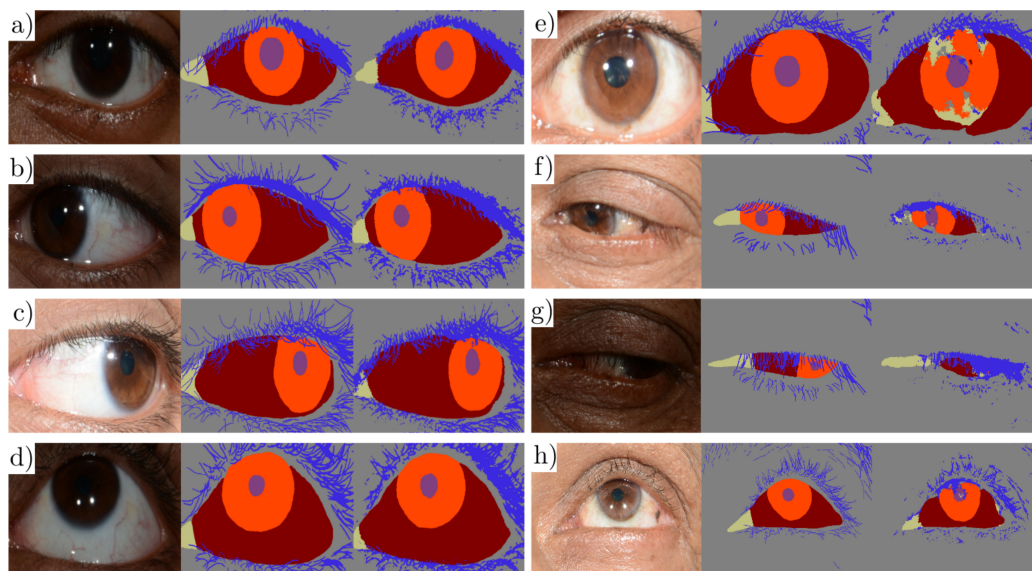
V tem podpoglavju analiziramo občutljivost SegNet modela glede na različne smeri pogleda. Zbirka MASD, ki smo jo uporabili v naših eksperimentih, vsebuje slike oči, ki gledajo v štiri različne smeri (pogled naravnost, levo, desno in navzgor). Želimo ugotoviti, ali se uspešnost naših modelov spreminja glede na smer pogleda. Ker MOS strategija zagotavlja razumne operativne točke, ki imajo primerljive vrednosti natančnosti in priklica, uporabimo strategijo analize občutljivosti. Rezultati naših eksperimentov so prikazani na sliki 6.6.



Slika 6.6: Analiza občutljivosti: grafi prikazujejo povprečno natančnost (levi graf zgoraj), priklic (desni graf zgoraj) ter F_1 -oceno (spodnji graf) glede na 4 različne smeri pogleda (naravnost, levo, desno ter navzgor). Standardne napake za povprečne vrednosti, ki jih beležimo zaradi 4-kratnega prečnega preverjanja, so prikazane na vrhu stolpcev.

Če primerjamo vrednosti priklica pri klasifikaciji zenice glede na smer pogleda, vidimo, da so vrednosti višje, ko oko gleda naravnost ali navzgor in nižje, kadar gleda levo ali desno. Najverjetneje je razlog za to dejstvo, da v podatkovni zbirki očesa, kjer oseba gleda levo ali desno, na slikah niso tako na široko odprta, kot pri pogledih naravnost/navzgor. Zaradi tega ob pogledih levo in desno regijo zenice prekrijejo trepalnice, ki povzročijo okluzije. Naslednji razlog je, da so vse šarenice v podatkovni zbirki MASD rjave barve. Kadar oko gleda levo/desno, se v kotičkih očesa šarenica obarva še temneje, zaradi česar model težko loči temno rjavo šarenico od črne zenice.

Segmentacija beločnice je v povprečju najbolj uspešna, kadar oko gleda navzgor (natančnost in priklic z MOS strategijo presežeta vrednost 0,95). Glavni razlog za to je po



Slika 6.7: Primeri najboljših in najslabših večrazrednih napovedi. Trojčki podslik a), b), c) in d) predstavljajo primere najboljših napovedi, trojčki e), f), g) in h) pa najslabših napovedi. V prvem stolpcu trojčkov so predstavljene originalne RGB slike, v drugem ročne anotacije teh slik ter v tretjem segmentacijski rezultat našega modela.

vsej verjetnosti dejstvo, da na tovrstnih slikah večinoma trepalnice ne zakrivajo oči. Drugi razlog je, da ob pogledu navzgor nastane le ena, neprekinjena regija beločnice, za razliko od drugih smeri pogledov, kjer šarenica vidno regijo beločnice tipično razdeli na dva dela. Ob pogledu levo/desno se pogosto pojavi majhna regija beločnice na strani medialnega ali lateralnega kantusa (odvisno od tega, v katero smer oko gleda), ki je pogosto nezaznana ali zamešana s samim kantusom.

Nekateri izmed najboljših in najslabših segmentacijskih rezultatov so za vsako smer pogleda prikazani na sliki 6.7. Na tej sliki prva vrstica slikovnih trojčkov prikazuje primere najboljših segmentacijskih rezultatov, druga vrstica trojčkov pa primere najslabših. Vsak slikovni trojček je sestavljen iz (gledano z leve proti desni): vhodne slike, ročne anotacije ter napovedi našega modela. Primer *a*) prikazuje pogled naravnost, *b*) pogled na levo, *c*) pogled na desno in *d*) pogled navzgor. Beločnica, zenica, šarenica in kantus so na teh primerih klasificirani (segmentirani) zelo natančno. Na primeru *a*) lahko opazimo, da je segmentirana zenica malenkost manjša, kakor na anotirani sliki. Na primerih *b*) in *c*) opazimo, da model detektira večino trepalnic, če so le-te locirane ena poleg druge, hkrati pa pogosto zgreši tiste, ki se na vhodni sliki pojavljajo ločeno od ostalih.

Na desni strani slike 6.7 so za vsako posamezno smer pogleda prikazani nekateri izmed najslabših segmentacijskih rezultatov. Na primeru *e*) lahko opazimo, da je velik predel

šarenice napačno klasificiran kot medialni kantus. Očesi, ki sta na slikah označeni kot *f*) in *g*) sta skoraj zaprti. Na sliki *f*) vidimo, da kantus skoraj ni zaznan, spodnji del šarenice je napačno segmentiran kot periokularna koža. Na sliki *g*) je prisotnih veliko okluzij zaradi trepalnic, ki prekrivajo šarenico, zaradi česar šarenica skoraj ni zaznana. Na sliki *h*) lahko opazimo, da manjši del desne strani beločnice ni zaznan zaradi temnih vzorcev v beločnici, ki se v uporabljeni podatkovni zbirki ne pojavljajo pogosto. Temen vzorec na beločnici je napačno klasificiran kot periokularna koža. Najvišji del šarenice je napačno klasificiran kot trepalnice (in delno kot periokularni del), kar je zelo verjetno posledica kontrastne tranzicije med šarenico in zgornjo obrvjo.

Tabela 6.2: Primerjava rezultatov segmentacije beločnice, ki jih dobimo z večrazrednim in enorazrednim segmentacijskim modelom SegNet.

Metrika uspešnosti	Natančnost	Priklic	F ₁ -ocena
Beločnica (enorazredni model)	0,88 ± 0,08	0,93 ± 0,07	0,90 ± 0,07
Beločnica (večrazredni model)	0,92 ± 0,04	0,90 ± 0,09	0,91 ± 0,07

6.6. Primerjava z enorazredno segmentacijo

V tem eksperimentu je narejena primerjava segmentacijske uspešnosti modela SegNet, ki je naučen na vseh šestih razredih, z uspešnostjo SegNet modela, ki je naučen le na enem razredu (je oziroma ni beločnica). Pri tem se bomo pri obeh modelih osredotočili izključno na primerjavo rezultatov segmentacije beločnice. Enorazredni model je učen z isto množico parametrov kakor večrazredni model, pri čemer sta opazovana časa učenja skoraj enaka. V tabeli 6.2 so navedeni rezultati, ki smo jih dobili z MOS strategijo.

Kot lahko vidimo iz predstavljenih rezultatov, sta modela podobno uspešna. Rezultati nakazujejo, da večrazredna segmentacija v primerjavi z enorazredno s stališča uspešnosti ni slabša, kljub temu, da je zmožna na vhodni sliki segmentirati več delov očesa. Poudariti je potrebno tudi, da je računska kompleksnost obeh modelov zelo podobna.

Kvalitativna primerjava segmentacijskih mask, ki jih pridobimo z obema modeloma (večrazrednim in enorazrednim), je predstavljena na sliki 6.8. Vidimo lahko, da oba modela večino slikovnih elementov napovesta enako. Največje razlike se pojavljajo pri delih, kjer se beločnica stika z drugimi deli očesa (meja s trepalnicami, kantusom itd.).



Slika 6.8: Kvalitativni rezultati segmentacije beločnice, ki jo pridobimo z večrazrednim in enorazrednim SegNet modelom po MOS strategiji. Prvi stolpec podslik (gledano z leve proti desni) prikazuje vhodne slike, drugi ročne anotacije beločnice, tretji rezultat večrazrednega modela in četrti rezultat enorazrednega modela.

6.7. Diskusija

Naši rezultati kažejo na to, da je večrazredni segmentacijski model zmožen slike očesa (v večje število ciljnih razredov) segmentirati z relativno visoko natančnostjo. Izvedli smo eksperimente, kjer primerjamo segmentacijo beločnice z večrazrednim in enorazrednim modelom, pri čemer se izkaže, da z obema segmentacijskima modeloma lahko dosegamo primerljive rezultate. Kljub temu, da je model sposoben klasificirati v več razredov, je računsko zahtevnost večrazrednega modela približno enaka enorazrednemu. Model za večrazredno segmentacijo bi v praksi lahko v marsikaterem fuzijskem cevovodu nadomestil množico modelov, ki segmentirajo le po eno modalnost. Taka nadomestitev ima torej velik potencial zmanjšati računsko zahtevnost obstoječih cevovodov.

Na tem mestu je smiselno poudariti, da so bili eksperimenti v tem poglavju narejeni na zbirki MASD in ne na zbirki SBVPI, ki jo uporabljamo v ostalih poglavjih tega dela. V našem fuzijskem sistemu kljub vzpodbudnim rezultatom, ki smo jih predstavili v tem poglavju, za segmentacijo ne bomo uporabljali tega pristopa. Glavni razlog za to je pomanjkanje večrazrednih anotacij, katerih izdelava je časovno zelo zahtevna. Pri zbirki

SBVPI namreč anotacij trepalnic in medialnega kantusa nimamo, hkrati pa smo velik del anotacij periokularnega dela in zenice pridobili šele v kasnejših fazah izdelave tega dela.

Poglavje 7

Prepoznavanje na podlagi periokularnega dela

7.1. Uvod

Periokularna regija je predel obraza, ki se nahaja v neposredni bližini očesa [56]. Razpoznavanje na podlagi te regije lahko interpretiramo kot nek kompromis med razpoznavanjem na podlagi obraza ter razpoznavanjem na podlagi očesnih modalnosti, pri čemer imata oba načina svoje prednosti in slabosti [113]. Šarenica tipično omogoča višjo natančnost razpoznavanja kot razpoznavanje na podlagi obrazov. Za pridobitev slik, ki so primerne za razpoznavanje na podlagi očesa, je ponavadi potrebno slediti strogemu protokolu: kamera mora biti dovolj blizu očesa, glava in pogled morata biti usmerjena v točno določeno smer, ne sme prihajati do premočnih odbojev svetlobe, veke ne smejo biti zaprte in v večini primerov je potrebna kooperacija subjektov. Po drugi strani je kamera pri zajemu obraza lahko bolj oddaljena, smer pogleda in pozicija sta precej manj kritični, od obraza ne prihaja do tako močnih odbojev svetlobe kot pri očesu in kooperacija subjektov večkrat ni potrebna. Ko sta šarenica ali beločnica ustrezno zajeti, je za razpoznavanje lahko uporabljena periokularna koža, ki vsebuje diskriminativno informacijo barve, gub, materinih znamenj ali oblike očesa, ki je vsebovana v periokularni regiji.

Definicija periokularne regije v literaturi vključuje enkrat več, drugič manj obočesne površine, kar lahko strnemo v dve različici definicije:

- *Ožja periokularna regija:* Poleg očesa vsebuje zgornjo in spodnjo veko in podočnjak. Slike, ki so na področju te regije zajete v dovolj visoki kvaliteti, vsebujejo tudi

informacijo o posameznih porah, manjših gubah in brazgotinah.

- *Širša periokularna regija:* Poleg ožje periokularne regije je na sliki prisotna še obrv, zgornji del lica in včasih del nosu. Pri slikah je tako manj poudarka na zelo finih kožnih detajlih (npr. porah), hkrati pa je vključena informacija o obliki obrvi, o prehodu kože iz očesnega dela na nosni del, kožnih strukturah na licu itd.

V tem poglavju se ukvarjamo z razpoznavanjem na podlagi ožje definicije periokularnega dela z metodami globokega učenja. Ožjo definicijo periokularne regije smo izbrali zato, ker je le-ta prisotna na slikah iz vaskularnega dela zbirke SBVPI (ki je namenjena raziskovanju beločnice), kar nam v nadaljevanju dela omogoča fuzijo vseh treh modalnosti (šarenica, beločnica in periokularni del) na istih vhodnih slikah.

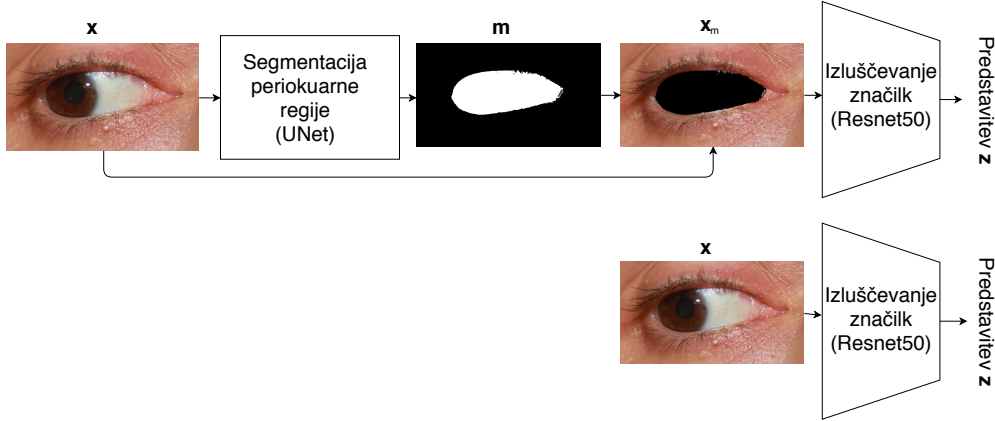
7.2. Metode

7.2.1. Pregled primerjanih cevovodov

V tem podpoglavju opišemo dva različna cevovoda za razpoznavanje na podlagi periokularnega dela ter primerjamo uspešnosti njunega razpoznavanja. Prvi cevovod, ki smo ga v naših eksperimentih testirali, je sestavljen iz dveh faz, in sicer: *i*) segmentacije periokularnega dela s pomočjo mreže UNet ter *ii*) izluščevanja značilk z ResNet50. Ta cevovod bomo zato v nadaljevanju poimenovali *UNet+ResNet*. Za primerjavo vzamemo cevovod, ki je predlagan v delu Deep-PRWIS [16]. Visokonivojski pregled, na kakšen način oba cevovoda iz vhodne slike \mathbf{x} pridobita numerično predstavitev \mathbf{z} , je prikazan na sliki 7.1. Zgornji cevovod na sliki prikazuje pristop z UNet+ResNet, spodnji pa Deep-PRWIS. Opazimo, da je razpoznavalni cevovod Deep-PRWIS enostavnejši od UNet+ResNet, saj v fazi izluščevanja značilk ne potrebuje procesa segmentacije. Glavna želja postopka Deep-PRWIS je namreč model naučiti, da diskriminativno informacijo išče le v periokularnem delu, informacijo iz šarenice in beločnice pa mora ignorirati. Ker informacije iz šarenice in beločnice model ne upošteva, segmentacija pri Deep-PRWIS v fazi izluščevanja ni potrebna. Učinek, da model informacije iz šarenice in beločnice ne upošteva, dosežemo s postopkom razširitve učne množice, ki je natančneje opisan v delu [16], opisali pa ga bomo tudi v nadaljevanju.

7.2.2. Opis cevovoda UNet+ResNet

Kot smo že omenili, visokonivojski pregled našega cevovoda za razpoznavanje na podlagi periokularnega dela sestavljata dve mreži, in sicer: *i*) UNet, ki pridobi regijo interesa (seg-



Slika 7.1: Prikaz dveh cevovodov za razpoznavanje na podlagi periokularnega dela, ki jih primerjamo v tem delu. Zgornji diagram prikazuje naš cevovod, pri katerem z uporabo mreže UNet iz vhodne slike \mathbf{x} najprej segmentiramo periokularni del in s pridobljeno binarno masko \mathbf{m} maskiramo vhodno sliko. Rezultat maskiranja \mathbf{x}_m podamo na vhod ResNet50, ki izlušči diskriminativne značilke. Spodnji diagram prikazuje postopek Deep-PRWIS, ki v fazi izluščevanja značilk ne potrebuje segmentacije, saj je model naučen tako, da ne upošteva informacije iz šarenice in beločnice.

mentacija periokularnega dela), ter *ii*) ResNet50, ki iz segmentiranega periokularnega dela izračuna numerično predstavitev slike. Shematičen prikaz cevovoda je prikazan na sliki 7.1 zgoraj. Sliko \mathbf{x} podamo na vhod segmentacijske mreže UNet, ki \mathbf{x} preslika v binarno masko \mathbf{m} . Če smo bolj natančni, mreža UNet dejansko na svoj izhod vrne napoved v obliki verjetnosti, končno binarno masko \mathbf{m} pa dobimo z upravljanjem teh verjetnosti glede na izbran prag $\Delta = 0,5$. Zaradi preprostejšega zapisa v nadaljevanju bomo binarizacijo šteli kot del preslikave, ki jo naredi UNet. Preslikavo nato lahko zapišemo kot:

$$\mathbf{m} = f_{\theta_1}(\mathbf{x}), \quad (7.1)$$

pri čemer θ_1 označuje parametre modela UNet, ki jih je potrebno naučiti. Z binarno masko \mathbf{m} maskiramo \mathbf{x} , s čimer izločimo slikovne elemente, ki ne pripadajo periokularnemu delu, kar lahko zapišemo kot:

$$\mathbf{x}_m = \mathbf{x} \odot \mathbf{m}, \quad (7.2)$$

pri čemer \odot predstavlja Hadamardov produkt. Maskirana slika \mathbf{x}_m gre nato na vhod mreže ResNet50, ki vrne izluščene značilke \mathbf{z} . Drugo preslikavo lahko označimo kot:

$$\mathbf{z} = f_{\theta_2}(\mathbf{x}_m), \quad (7.3)$$

pri čemer θ_2 označuje parametre modela ResNet50, ki jih moramo naučiti.

7.2.3. Opis cevovoda Deep-PRWIS

Glavna hipoteza v delu Deep-PRWIS [16] je, da z ignoriranjem informacije iz šarenice in beločnice lahko izboljšamo natančnost razpoznavanja na podlagi periokularne regije. Eden izmed prispevkov tega dela je predlagan način za učenje modela, s katerim dosežemo, da se model samostojno nauči ignorirati informacijo iz šarenice in beločnice. Zaradi tega segmentacije v času napovedovanja ne potrebujemo. Pri naši implementaciji Deep-PRWIS se članka ne držimo dosledno, ampak iz njega vzememo bistvene ideje za učenje modela (arhitekturo za izluščevanje značilnk smo namreč nadomestili z drugo). V izvornem delu za izluščevanje značilnk uporabljajo plitvejšo arhitekturo, mi pa smo jo zamenjali z ResNet50 (za ta del cevovoda lahko uporabimo katero koli razpoznavalno CNN arhitekturo). Naslednji vzrok za zamenjavo arhitekture pri konkurenčnem algoritmu je dejstvo, da v izvornem delu avtorji uporabljajo plitvejšo arhitekturo, ki jo učijo na bolj obširnih podatkovnih zbirkah (UBIRIS.v2 in FRGC). Ker je naša podatkovna zbirka SBVPI od omenjenih zbirk precej manjša, v predhodnih eksperimentih na njihovi arhitekturi nismo mogli naučiti uspešnega modela. Namesto tega, da bi vse parametre njihove arhitekture učili sami od začetka (pri čemer kode oz. modelov nismo uspeli dobiti), smo se (kakor je opisano v nadaljevanju) poslužili prenešenega učenja z uporabo arhitekture ResNet50 in inicializacijo parametrov ImageNet [92]. Kot je razvidno na sliki 7.1 (spodnji cevovod), na vhod cevovoda Deep-PRWIS damo celotno, nesegmentirano sliko očesa \mathbf{x} . Preslikavo vhodne slike \mathbf{x} v numerično predstavitev \mathbf{z} , lahko opišemo kot:

$$\mathbf{z} = f_{\theta_1}(\mathbf{x}), \tag{7.4}$$

pri čemer so θ_1 parametri modela ResNet50, ki jih moramo naučiti.

Čeprav segmentacije v fazi napovedovanja ne potrebujemo, je na tem mestu potrebno poudariti, da segmentacijski model, ki je zmožen ločiti periokularen del od očesnega, kljub temu potrebujemo za pripravo učnih podatkov. Ko je model enkrat naučen, pa le-tega ne potrebujemo več. Postopek za učenje modela Deep-PRWIS bo natančneje predstavljen v naslednjem podpoglavju.

7.3. Učenje modelov in generiranje izhoda

7.3.1. Učenje segmentacijskega modela UNet

Segmentacijski model UNet potrebujemo za pripravo učnih podatkov pri obeh cevovodih, v fazi napovedovanja pa ga potrebujemo le pri modelu UNet+ResNet. Za učenje parametrov segmentacijskega modela smo slike iz zbirke SBVPI razdelili na dve disjunktne množici:

- *Učna množica* vsebuje 510 slik s korespondenčnimi, ročno anotiranimi maskami periokularnega dela. Parametri modela so učeni na 460 slikah, ostalih 50 slik pa je v fazi učenja uporabljenih za spremljanje poteka učenja (validacijska množica).
- *Testna množica* vsebuje 1.312 slik z ročno izdelanimi korespondenčnimi anotacijami.

V testni množici je občutno več slik kot v učni množici, ker smo za učenje segmentacijskega modela uporabili vse anotacije, ki smo jih imeli tistem času na voljo. To nam je omogočilo nadaljnji razvoj cevovoda, segmentacijski model pa smo ovrednotili šele, ko smo uspeli označiti vse slike v testni množici. Za učenje parametrov segmentacijskega modela f_{θ_1} smo slike učne množice zmanjšali na standardno velikost 320×320 slikovnih elementov. Ker je učna množica relativno velika, je nismo širili z generiranjem novih slik, kakor smo to delali pri segmentaciji beločnice. Model učimo z Adam optimizacijo, pri čemer nastavimo stopnjo učenja na 0,0001. Za funkcijo izgube izberemo binarno prečno entropijo (angl. binary cross entropy). Velikost paketov (angl. batch size) nastavimo na 15 in model učimo, dokler njegova natančnost ne konvergira, kar se v naših eksperimentih zgodi po 16-ih dobah.

7.3.2. Učenje modela ResNet50 pri UNet+ResNet

Razdelitev zbirke SBVPI, ki jo uporabljamo za eksperimente razpoznavanja na podlagi periokularnega dela, je enaka razdelitvi, ki smo jo opisali pri razpoznavanju na podlagi beločnice v podpoglavju 5.3.2. 220 identitet je na enak način razdeljenih v učno in testno množico. Enaka razdelitev je v obeh primerih narejena zato, ker bomo to lastnost potrebovali v poglavju, kjer se ukvarjamo s fuzijo. Kasneje bomo videli, da tudi tretji razpoznavalni cevovod na podlagi šarenice uporablja enako razdelitev, zaradi česar so vsi trije cevovodi primerljivi glede na uspešnost razpoznavanja, obenem pa lahko merimo uspešnost fuzije modalnosti. Učno množico obogatimo z naključnimi rotacijami (do ± 20

stopinj) in naključnimi afinimi transformacijami, pri čemer iz vsake izvorne slike izdelamo 8 novih slik.

Uporabili smo implementacijo arhitekture ResNet50, ki je dostopna v knjižnici Keras. Predzadnji sloj arhitekture je polno-povezani sloj velikosti 2.048. Zadnji sloj je *softmax*, ki klasificira v 1.000 razredov, saj je tako definiran problem tekmovanja ImageNet [92]. Za našo uporabo smo zadnja dva sloja odstranili in ju nadomestili s slojem globalnega povprečnega združevanja, ki mu sledi polno-povezani sloj s 512 nevroni ter na koncu še *softmax*, ki klasificira v naših 120 učnih identitet. Model učimo s prenešenim učenjem, kar pomeni da pred učenjem inicializiramo uteži na vrednosti, ki so se dobro izkazale na drugi domeni (v našem primeru gre za tekmovanje ImageNet). Nato učimo le zadnje tri sloje, ki smo jih sami dodali arhitekturi.

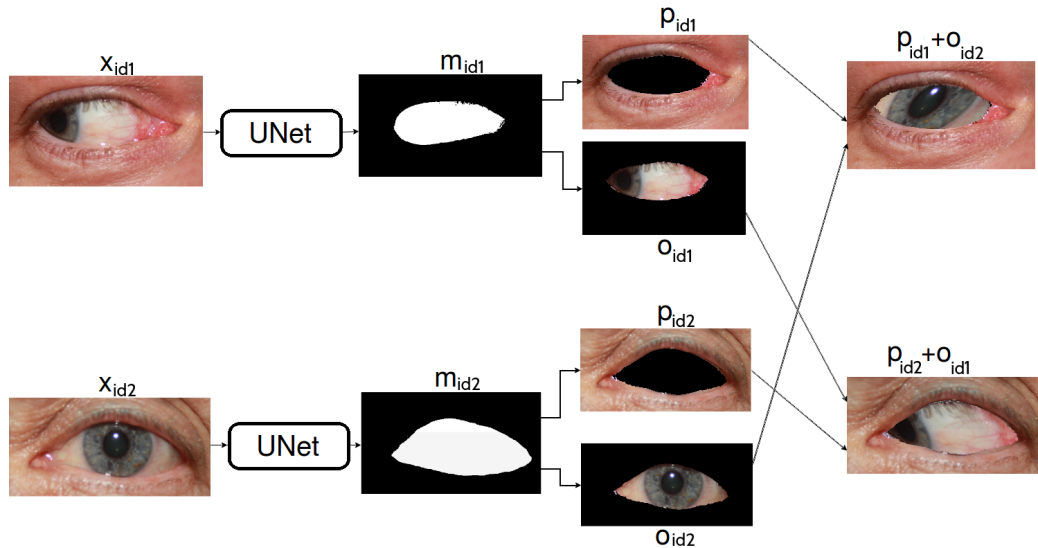
Za optimizacijo parametrov uporabimo Adam optimizator s stopnjo učenja 0,0001, kot funkcijo izgube pa uporabimo kategorično prečno entropijo. Rezultati na validacijski množici začnejo konvergirati po 16 dobah.

7.3.3. Učenje modela ResNet50 pri Deep-PRWIS

Razdelitev zbirke SBVPI na učno in testno množico je ponovno konsistentna s splošno uporabljenim protokolom v tem delu. Učno množico obogatimo po postopku, ki je opisan v delu [16] in prikazan na sliki 7.2. Nov učni primer s slike \mathbf{x}_{id1} , ki pripada identiteti $id1$ dobimo tako, da si najprej izberemo poljubno sliko \mathbf{x}_{id2} , ki pripada drugi identiteti $id2$. Z UNet modelom iz \mathbf{x}_{id1} in \mathbf{x}_{id2} izračunamo korespondenčni segmentacijski maski \mathbf{m}_{id1} in \mathbf{m}_{id2} . S slike \mathbf{x}_{id1} in maske \mathbf{m}_{id1} (oziroma njene negacije) lahko z Hadamardovim produktom izločimo slikovne elemente očesnega dela (oziroma obočesnega dela). Na ta način pridobimo sliko periokularnega dela \mathbf{p}_{id1} in sliko okularnega dela \mathbf{o}_{id1} . Na enak način iz \mathbf{x}_{id2} in \mathbf{m}_{id2} izdelamo sliko \mathbf{p}_{id2} in \mathbf{o}_{id2} . Nov učni primer za identiteto $id1$ dobimo z združevanjem obočesnega dela \mathbf{p}_{id1} in očesnega dela \mathbf{o}_{id2} , kar bi lahko označili kot $\mathbf{p}_{id1} + \mathbf{o}_{id2}$ (kjer $+$ ne predstavlja klasičnega seštevanja, vendar pomeni le združevanje delov). Preden sliko dejansko združimo, moramo sliko očesa \mathbf{o}_{id2} s homografsko transformacijo preoblikovati na tak način, da se bo njegova zunanja oblika v čim večji meri ujemala z obliko črnine, ki je zaradi izreza očesnega dela nastala na sliki periokularnega dela \mathbf{p}_{id1} .

Vsaka slika v obogateni učni množici torej vsebuje periokularni in okularni del (za razliko od našega modela, kjer okularnega dela nima nobena slika, ne v učni, ne v testni množici). Model se v fazi učenja zato nauči, da se iz okularnega dela na sliki ne da sklepati na identiteto. Za napoved posledično upošteva le značilke iz periokularnega dela, kar pomeni, da eksplicitna segmentacija v cevovodu ni več potrebna. Za učenje ponovno uporabimo Keras implementacijo arhitekture ResNet50, ki jo spremenimo oz. dopolnimo

na identičen način, kot smo to naredili pri cevovodu *UNet + ResNet*. Model potrebuje 10 dob, da njegova uspešnost na validacijski množici začne konvergirati.



Slika 7.2: Prikaz Deep-PRWIS postopka za obogatitev učne množice SBVPI, pri katerem je cilj združiti periokularni del identitete $id1$ ter okularni del identitete $id2$ (ali obratno). S segmentacijskim modelom UNet iz slik x_{id1} in x_{id2} naredimo korespondenčne maske m_{id1} in m_{id2} . S pomočjo mask na vhodnih slikah maskiramo periokularno in okularno regijo. Za identiteto $id1$ lahko naredimo nov učni primer tako, da združimo periokularno regijo p_{id1} ter okularno regijo o_{id2} . Pred združitvijo s homografsko transformacijo preoblikujemo obliko okularne regije tako, da se ujema z obliko črnine, ki je zaradi izreza okularne regije prisotna na sliki p_{id1} .

7.3.4. Metodologija razpoznavanja

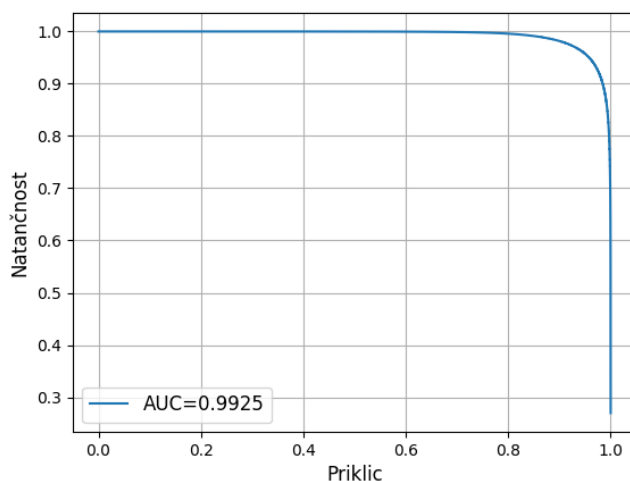
Ko sta oba razpoznavalna modela ResNet50 (eden pripada cevovodu UNet+ResNet, drugi pa Deep-PRWIS) naučena, ju lahko uporabimo za luščenje značilk. Obema ResNet50 arhitekturama moramo odstraniti *softmax* sloj, značilke, ki opisujejo vhodno sliko, pa vzamemo iz združevalnega sloja globalnega povprečja. Gre za podoben postopek preoblikovanja CNN arhitekture (katerega cilj je uporaba CNN modela kot izluščevalca značilk), kot smo ga uporabili pri mreži ScleraNET v poglavju 5.2.4. To preoblikovanje nam omogoči izračun ujemanja med primeri, ki jih razpoznavalna modela v fazi učenja še nista videla. Ujemanje med parom vhodnih slik dobimo tako, da izračunamo kosinusno podobnost med

vektorjema značilnk, ki ju vrne posamezen ResNet50 model.

7.4. Eksperimenti in rezultati

7.4.1. Ovrednotenje segmentacije

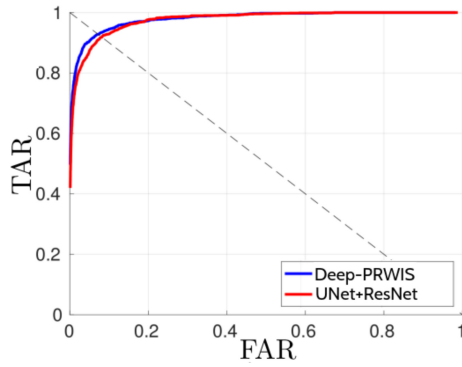
Na sliki 7.3 je prikazana krivulja natančnost/priklic, ki prikazuje uspešnost segmentacijskega modela UNet na testni množici. Pri upragovanju z $\Delta = 0,5$ dobimo natančnost 0,975, priklic 0,919 ter F_1 -oceno 0,946.



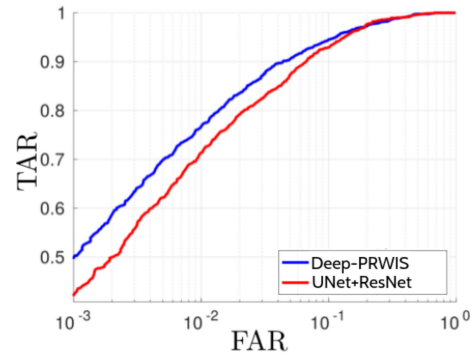
Slika 7.3: Krivulja natančnost/priklic, ki prikazuje uspešnost UNet modela za segmentacijo periokularne regije.

7.4.2. Ovrednotenje razpoznavanja

V tabeli 7.1 so navedeni rezultati verifikacijskih in identifikacijskih eksperimentov. Celotna ROC krivulja verifikacijskih eksperimentov je prikazana na sliki 7.4, CMC krivulja identifikacijskih eksperimentov pa na sliki 7.5. Opazimo, da oba cevovoda zagotavljata precej podobno uspešnost, pri čemer je Deep-PRWIS pristop na vseh grafihi uspešnejši.

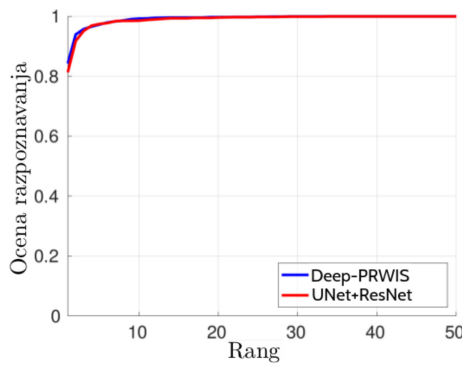


(a) Na linearni skali prikazani ROC krivulji.

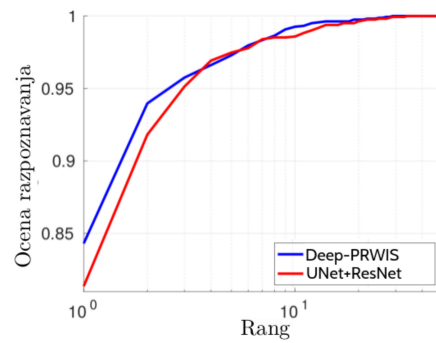


(b) Na semi-logaritmični skali prikazani ROC krivulji.

Slika 7.4: Rezultati verifikacijskih eksperimentov, v katerih primerjamo uspešnost cevovodov UNet+ResNet in Deep-PRWIS.



(a) Na linearni skali prikazani CMC krivulji.



(b) Na semi-logaritmični skali prikazani CMC krivulji.

Slika 7.5: Rezultati identifikacijskih eksperimentov, v katerih primerjamo uspešnost cevovodov UNet+ResNet in Deep-PRWIS.

Tabela 7.1: Rezultati eksperimentov razpoznavanja na podlagi periokularnega dela. Rezultati so ločeno prikazani za verifikacijske in identifikacijske eksperimente.

Eksperiment	Algoritem	VER@0.1FAR	VER@1FAR	EER	AUC
Verifikacija	UNet+ResNet	0,424	0,713	0,082	0,959
	Deep-PRWIS	0,499	0,768	0,077	0,961
	Algoritem		Rank-1	Rank-5	AUCMC
Identifikacija	UNet+ResNet		0,814	0,975	0,991
	Deep-PRWIS		0,843	0,973	0,993

7.5. Diskusija

V tem poglavju smo preizkusili in ovrednotili dva cevovoda za razpoznavanje periokularne regije na podlagi globokega učenja. Ugotovili smo, da oba pristopa zagotavljata relativno podobno uspešnost razpoznavanja, pri čemer Deep-PRWIS v vseh eksperimentih nekoliko presega uspešnost cevovoda UNet+ResNet. S pristopom Deep-PRWIS lahko zmanjšamo kompleksnost razpoznavalnega cevovoda, saj segmentacija v fazi napovedovanja ni potrebna. Obenem nam pristop zagotovi nekoliko višje natančnosti razpoznavanja, tako pri verifikacijskih kot pri identifikacijskih eksperimentih. Cena teh dveh prednosti je zahtevnejši postopek priprave učnih podatkov, pri katerem moramo učno množico obogatiti z generiranjem novih, umetnih primerov. Pri tem ohranimo periokularni del izbrane identitete, okularni del pa vzamemo od drugih identitet. Zaradi predstavljenih prednosti cevovoda Deep-PRWIS ta cevovod porabimo tudi v fuziji modalnosti v nadaljevanju dela.

Poglavje 8

Prepoznavanje na podlagi šarenice

8.1. Uvod

V tem poglavju je predstavljen cevovod za razpoznavanje na podlagi šarenice. Razvoj dotičnega cevovoda ni del tega magistrskega dela, ampak je del magistrske naloge [114], katere avtor je Juš Lozej. V naši nalogi bo uporabljen za pridobitev referenčnih števil, s pomočjo katerih bomo lahko ovrednotili predhodno obravnavana cevovoda (na podlagi beločnice in periokularnega dela). Cevovod bo uporabljen tudi v naslednjem poglavju, v katerem bomo vse tri sisteme združili v fuzijo biometričnih modalnosti. Cevovod so prvotno razvijali na podatkovni zbirki CASIA Iris Thousand [62], ki je namenjena predvsem razpoznavanju na podlagi šarenice, v tem delu pa ga bomo ovrednotili na zbirki SBVPI. Zbirka CASIA, za razliko od SBVPI (pri kateri imamo 4 različne poglede za isto identiteto), vsebuje slike, na katerih večina očes gleda naravnost, zato ta kriterij naredi SBVPI zbirko bolj zahtevno. Cevovod je v celoti osnovan na metodah globokega učenja. Tako kot prej predstavljene cevovode ga bomo ovrednotili na primerih, ki jih sistem še ni videl v fazi učenja (angl. open-world setting).

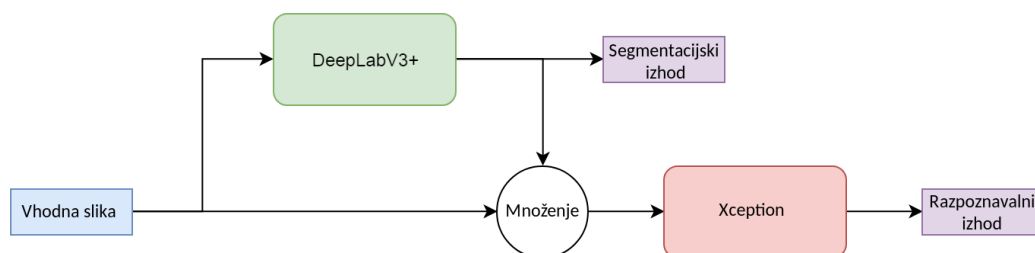
8.2. Metode

Arhitektura razpoznavalnega cevovoda na podlagi šarenice, ki je prikazana na sliki 8.1, je sestavljena iz dveh višjenivojskih gradnikov, in sicer *i*) segmentacijskega dela ter *ii*) razpoznavalnega dela. Segmentacijski del cevovoda predstavlja arhitektura DeepLabV3+, razpoznavalni del pa arhitektura Xception. Vhodna slika gre v segmentacijski del, ki iz nje segmentira šarenico. Vhodne slike nato množimo s segmentacijskim izhodom, s čimer izločimo slikovne elemente na sliki, ki ne pripadajo šarenici. Segmentirana šarenica gre naprej v Xception, ki je razpoznavalni del cevovoda. Parametre modela smo učili v treh ločenih korakih, in sicer: *i*) učenje DeepLabV3+, *ii*) učenje Xception in *iii*) natančno nastavljanje parametrov Xception modula.

V prvem koraku smo s slikami iz SBVPI (in korespondenčnimi anotacijami šarenice) ločeno učili samo segmentacijski del mreže, torej DeepLabV3+, preostalega cevovoda pa v tej fazi nismo uporabljali. Učne primere smo pomanjšali na velikost 320×320 slikovnih elementov. Za optimizacijo parametrov smo uporabljali Adam optimizator s stopnjo učenja 0,0001 in cenilno funkcijo kategorične prečne entropije. Učna množica segmentacijskega modela je vsebovala 332 primerov, pri čemer smo jih 274 uporabili za dejansko učenje parametrov, 58 pa kot validacijsko množico.

V drugem učnem koraku segmentacijski arhitekturi dodamo preostanek celotne arhitekture, torej množenje vhodne slike s segmentacijskim izhodom ter Xception modul. Učenje segmentacijskega modela tokrat izklopimo – z drugimi besedami, učimo le Xception, v DeepLabV3+ pa se uteži, ki smo jih optimizirali v prejšnjem koraku, med učenjem ne posodablja več. Xception ima v fazi učenja na koncu svoje arhitekture *softmax* sloj, ki klasificira v naših 120 učnih razredov. V testni fazi Xception modulu odstranimo zadnji *softmax* sloj in izluščene značilke (enodimenzionalni vektor dolžine 2.048) vzamemo iz sloja globalnega povprečnega združevanja. Končno podobnost med identitetami (ki jih model v fazi učenja še ni videl) dobimo z izračunom kosinusne podobnosti med vektorji značilk. Kot učne podatke za razpoznavalni del cevovoda smo uporabili enake podatke kot pri cevovodih za razpoznavanje na podlagi beločnice in periokularnega dela. Za optimizacijo parametrov smo ponovno uporabili Adam optimizator s stopnjo učenja 0,0001 in model učili, dokler uspešnost razpoznavanja ni začela konvergirati.

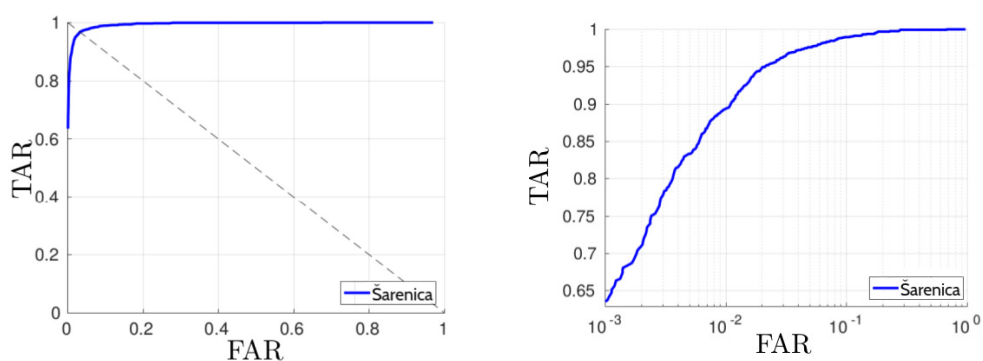
V tretji učni fazi smo zmanjšali stopnjo učenja na 0,000001, s čimer smo nastavljali natančne podrobnosti modela (angl. fine-tuning).



Slika 8.1: Prikaz arhitekture za razpoznavanje na podlagi šarenice [114]. Vhodna slika gre v segmentacijski model, ki segmentira šarenico. Segmentacijski izhod pomnožimo z vhodno sliko, s čimer izločimo slikovne elemente, ki ne pripadajo šarenici. Segmentirana šarenica gre naprej v razpoznavalni del cevovoda.

8.3. Rezultati

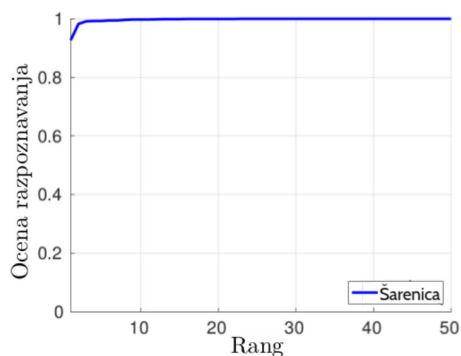
Rezultati verifikacijskih eksperimentov so prikazani na sliki 8.2, identifikacijskih eksperimentov pa na sliki 8.3. Rezultate za izbrane točke ter ploščine pod krivuljami beležimo v tabeli 8.1.



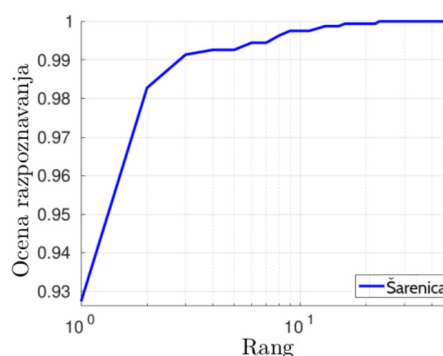
(a) Na linearni skali prikazana ROC krivulja.

(b) Na semi-logaritmični skali prikazana ROC krivulja.

Slika 8.2: Rezultati verifikacijskih eksperimentov s cevovodom za razpoznavanje na podlagi šarenice.



(a) Na linearni skali prikazana CMC krivulja.



(b) Na semi-logaritmični skali prikazana CMC krivulja.

Slika 8.3: Rezultati identifikacijskih eksperimentov s cevovodom za razpoznavanje na podlagi šarenice.

Tabela 8.1: Rezultati razpoznavalnih eksperimentov. Rezultati so ločeno prikazani za verifikacijske in identifikacijske eksperimente.

Ekspersiment	VER@0.1FAR	VER@1FAR	EER	AUC
Verifikacija	0,636	0,894	0,033	0,964
		Rank-1	Rank-5	AUCMC
Identifikacija		0,927	0,993	0,998

8.4. Diskusija

Šarenica se uvršča med biometrične modalnosti, na podlagi katerih je mogoče razpoznavati z izjemno visoko stopnjo natančnosti. V primerjavi s cevovodi, ki smo jih obravnavali v prejšnjih poglavjih, ima razpoznavanje na podlagi šarenice najvišjo natančnost razpoznavanja (tako pri verifikacijskih kot pri identifikacijskih eksperimentih). SBVPI ima v primerjavi z drugimi zbirkami, ki so namenjene izključno razpoznavanju na podlagi šarenice, dodatno dimenzijo težavnosti, saj SBVPI za isto identiteto vsebuje slike 4 različnih pogledov. Ko cevovod primerja dve očesi iste identitete, ki ne gledata v isto smer (npr. eno gleda levo in drugo desno), se presek vidnih delov šarenice, ki ju dobimo v obeh primerih, lahko precej zmajša. Ob pogledih levo in desno imamo težavo s tem, da se del šarenice, ki

sovpada s smerjo pogleda (ob pogledu levo, gre za levi del šarenice, ob pogledu desno pa za desni del šarenice), optično pomanjša. Očesno zrklo je namreč okroglo, kar povzroči, da je del šarenice, ki je nasproten smeri pogleda, bližje kameri, del šarenice, ki sovpada s smerjo pogleda, pa je od kamere bolj oddaljen. Veke pogosto prekrijejo predel šarenice, kateri del bo prekrit, pa je ponovno odvisno od smeri pogleda.

Poglavje 9

Fuzija modalnosti

9.1. Uvod

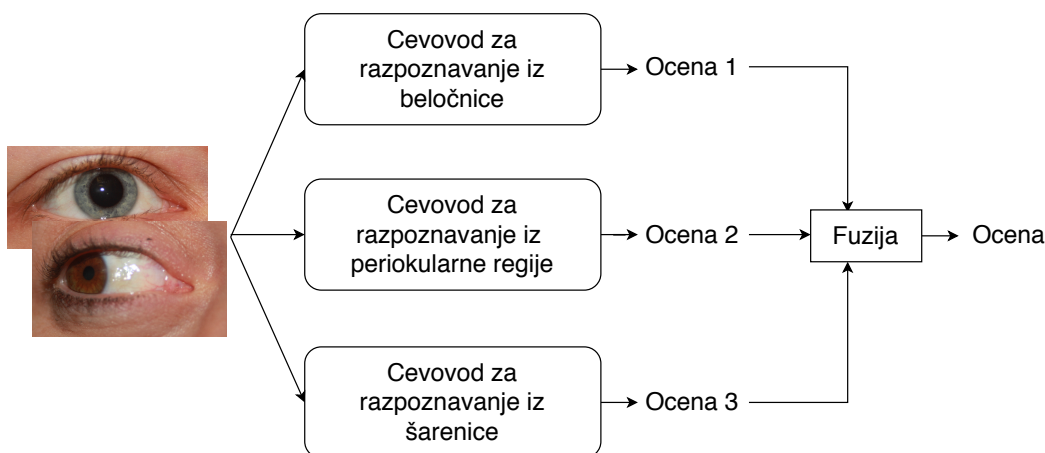
V prejšnjih poglavjih smo razvili tri cevovode, pri čemer vsak cevovod razpozna na podlagi druge biometrične modalnosti. V tem poglavju vse cevovode združimo v fuzijski sistem. Osnovna ideja biometrične fuzije je, da za pridobitev končne napovedi razpoznavnega sistema hkrati uporabimo informacijo iz večjega števila biometričnih modalnosti. S fuzijo biometričnih modalnosti lahko dosežemo višjo natančnost biometričnega sistema kot jo zagotavlja biometričen sistem, ki deluje le na podlagi posamezne modalnosti [12].

9.2. Metode

Poznamo več vrst biometrične fuzije, glede na to, na katerem nivoju razpoznavnega sistema združujemo informacijo različnih modalnosti, in sicer [115]:

- Pred izračunom ujemanja (angl. before matching).
 - Na nivoju senzorja (angl. sensor level).
 - Na nivoju značilk (angl. feature level).
- Po izračunu ujemanja (angl. after matching).
 - Na nivoju ocene (angl. score level).
 - Na nivoju ranga (angl. rank level).
 - Na nivoju odločitve (angl. decision level).

V vseh naših eksperimentih smo uporabljali izključno fuzijo na nivoju ocene. Ocena v



Slika 9.1: Shematičen prikaz naše fuzije na nivoju ocene. Predstavljajmo si, da nas zanima fuzijska ocena podobnosti med parom vhodnih slik. Vhodni sliki gresta ločeno v tri različne razpoznavalne cevovode. Vsak posamezen cevovod na izhod vrne oceno podobnosti med vhodnim parom slik, vse ocene pa so preko fuzijske operacije združene v končno oceno podobnosti.

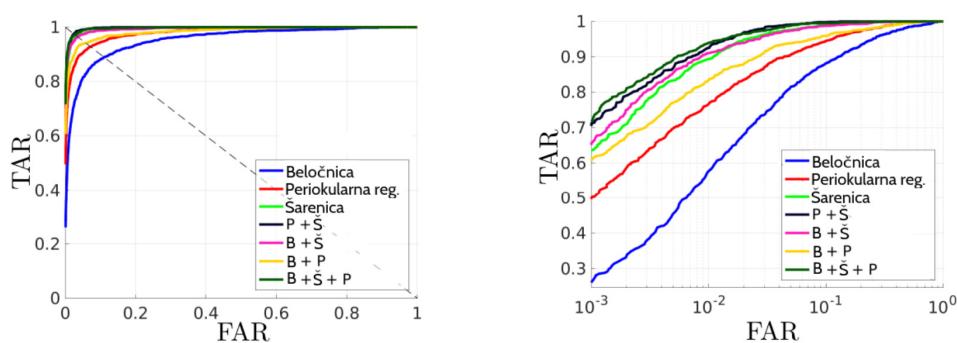
našem primeru konkretno pomeni izračunano podobnost oz. ujemanje med posameznimi pari slik oči. Na sliki 9.1 je shematičen prikaz fuzije treh razpoznavalnih cevovodov na nivoju ocene, ki smo jo naredili v tem delu. Fuzijsko oceno med parom oči dobimo tako, da par oči ločeno podamo na vhod naših treh cevovodov, ki razpoznavajo na podlagi različnih očesnih modalnosti. Za razpoznavanje na podlagi beločnice uporabljamo cevovod z dvofazno pridobitvijo regije interesa in ScleraNET za izluščevanje značilnk. Za razpoznavanje na podlagi periokularnega dela uporabimo Deep-PRWIS pristop, pri čemer uporabimo CNN arhitekturo ResNet50. Za razpoznavanje na podlagi šarenice pa uporabimo cevovod, ki je bil razvit v sklopu magistrskega dela, katerega avtor je Juš Lozej [114]. Vsak izmed teh treh cevovodov posebej za izbran par vhodnih slik vrne oceno podobnosti med slikama. Ocene podobnosti med vhodnimi pari slik pri posameznem cevovodu dobimo tako, da s kosinusno razdaljo merimo podobnost vektorjev značilnk, ki jih pridobimo iz mreže v cevovodu, ki služi kot izluščevalec značilnk. Vse tri ocene preko fuzijske operacije združimo v skupno fuzijsko oceno. V našem delu smo ovrednotili dve različni fuzijski operaciji, in sicer vsoto ter produkt. Končna fuzijska ocena je preprosto seštevek oziroma produkt ocen posameznih cevovodov.

9.3. Rezultati

V tabeli 9.1 so prikazani rezultati različnih vrst fuzije. Na začetku tabele imamo (zaradi referenčnih vrednosti) navedene rezultate razpoznavalnih cevovodov brez fuzije. Cevovodi so v tabeli označeni s prvo črko modalnosti, na podlagi katere delujejo (B – cevovod na podlagi beločnice, P – cevovod na podlagi periokularne regije, Š – cevovod na podlagi šarenice). Nato sledijo rezultati vseh možnih parov fuzij, ki jih lahko naredimo med tremi različnimi cevovodi. Za vsak par predstavimo rezultate dveh različnih fuzijskih operacij, in sicer za operacijo seštevanja (v tabeli označeno +) in množenja (označeno kot *). Zadnji dve vrstici prikazujeta fuzijo vseh treh razpoznavalnih cevovodov.

Pri verifikacijskih eksperimentih v tabeli lahko opazimo, da ima fuzija s seštevanjem (v primerjavi s fuzijo z množenjem) konsistentno višjo vrednost AUC. Za eksperimente identifikacije se izkaže, da izbira fuzijske operacije na uspešnost sistema vpliva mnogo manj. Zaradi teh lastnosti za naš fuzijski sistem raje izberemo operacijo seštevanja.

Na sliki 9.2 so izrisane verifikacijske ROC krivulje, ki prikazujejo uspešnost različnih seštevalnih kombinacij, ki jih dobimo s fuzijo modalnosti. Opazimo, da z večanjem števila združenih modalnosti, konsistentno dobivamo boljše verifikacijske rezultate. Podobno velja za identifikacijske eksperimente, za katere so na sliki 9.3 izrisane CMC krivulje.



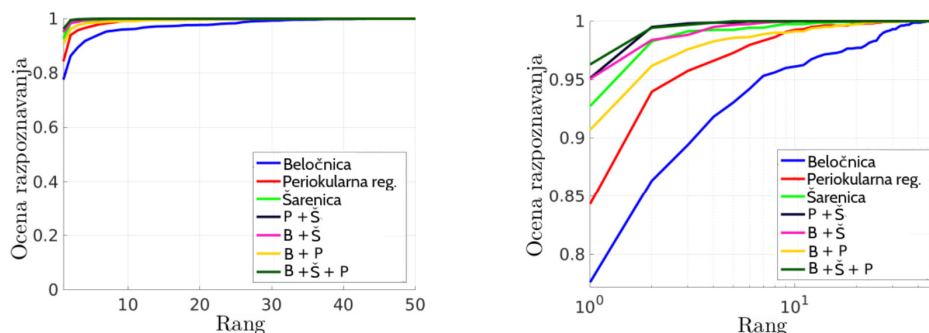
(a) Na linearni skali prikazane ROC krivulje.

(b) Na semi-logaritmični skali prikazane ROC krivulje.

Slika 9.2: Rezultati fuzijskih verifikacijskih eksperimentov. Na levem grafu so ROC krivulje prikazane na linearni skali, medtem ko desni graf prikazuje ROC krivulje na semi-logaritmični skali, zaradi česar je bolj vidna razlika pri nižjih FAR vrednostih.

Tabela 9.1: Rezultati fuzije, pri čemer velika črka označuje prisotnost cevovoda na podlagi določene fuzije v fuzijskem sistemu (B je beločnica, P periokularna regija, Š pa šarenica), operacija med črkami pa fuzijsko operacijo med ocenami (+ seštevanje, * množenje).

Eksperiment	Algoritem	VER@0.1FAR	VER@1FAR	EER	AUC
Verifikacija	B	0,261	0,579	0,111	0,947
	P	0,499	0,768	0,072	0,961
	Š	0,636	0,894	0,033	0,964
	P+Š	0,708	0,926	0,024	0,996
	P*Š	0,702	0,924	0,026	0,627
	B+Š	0,655	0,910	0,036	0,993
	B*Š	0,656	0,911	0,037	0,868
	B+P	0,596	0,835	0,058	0,983
	B*P	0,606	0,831	0,057	0,952
	B+P+Š	0,721	0,940	0,027	0,996
	B*P*Š	0,715	0,936	0,028	0,846
	Algoritem		Rank-1	Rank-5	AUCMC
Identifikacija	B		0,776	0,931	0,976
	P		0,843	0,973	0,993
	Š		0,927	0,993	0,998
	P+Š		0,951	0,999	0,999
	P*Š		0,953	0,999	0,999
	B+Š		0,951	0,997	0,999
	B*Š		0,948	0,996	0,999
	B+P		0,907	0,986	0,995
	B*P		0,906	0,985	0,995
	B+P+Š		0,963	1	0,999
	B*P*Š		0,959	0,999	0,999



(a) Na linearni skali prikazane CMC krivulje.

(b) Na semi-logaritmični skali prikazane CMC krivulje.

Slika 9.3: Rezultati fuzijskih identifikacijskih eksperimentov. Na levem grafu so CMC krivulje prikazane na linearni skali, medtem ko desni graf prikazuje CMC krivulje na semi-logaritmični skali, zaradi česar je bolj vidna razlika pri nižjih rangih.

9.4. Diskusija

V tem poglavju smo ovrednotili različne pristope za fuzijo treh očesnih modalnosti, pri čemer smo fuzijo izvajali na nivoju ocene. Naredili smo primerjavo med uspešnostjo sistemov, ki razpoznavanje na podlagi ene modalnosti, med fuzijskimi kombinacijami, pri čemer združujemo po dve biometrični modalnosti ter med skupno fuzijo vseh treh modalnosti. Pri tem smo ovrednotili tudi dve različni fuzijski operaciji, in sicer seštevanje in množenje.

Kot smo pričakovali, ima v obeh razpoznavalnih scenarijih (pri identifikaciji in verifikaciji) najvišjo natančnost razpoznavanje na podlagi šarenice. Sledi razpoznavanje na podlagi periokularnega dela, nato pa razpoznavanje na podlagi vaskularnih struktur iz beločnice. To, da je bila uspešnost razpoznavanja periokularnega dela višja kakor razpoznavanje na podlagi beločnice, je po vsej verjetnosti posledica dveh dejavnikov: *i*) neodvisnost periokularne regije od smeri pogleda in *ii*) diskriminativnost periokularne regije. Strukture na periokularnem delu so neodvisne od smeri pogleda, pri čemer vidne strukture žil v beločnici močno spreminjajo svojo obliko že ob manjših premikih očesa, saj prihaja do kompleksnih nelinearnih deformacij. Ožji periokularni del nosi veliko različnih tipov diskriminativne informacije, npr. oblika očesa, barva kože, gube, pore, materina znamenja, iz katerih lahko sklepamo tudi na nekatere mehke modalnosti (npr. spol, etnična pripadnost itd). V vaskularnih strukturah je tovrstnih tipov informacije veliko manj, saj se zanašamo le na vzorce zajetih struktur.

V naših rezultatih se izkaže, da se natančnost fuzijskega sistema konsistentno povečuje s povečevanjem števila vključenih modalnosti. V vseh primerih, ko obstoječemu razpoznavalnemu cevovodu s fuzijo dodamo novo modalnost, dobimo višjo natančnost novega fuzijskega sistema, kakor je bila natančnost obstoječega sistema. Dodajanje nove modalnosti v fuzijo v nobenem izmed naših eksperimentov ni poslabšalo prejšnje uspešnosti. V naših eksperimentih se je izkazalo, da v verifikacijskem načinu združevanje s fuzijsko operacijo seštevanja deluje boljše, kakor s fuzijsko operacijo množenja. Ker v identifikacijskem načinu obe operaciji zagotavljata podobne rezultate, smo se odločili, da za naš končni sistem uporabimo operacijo seštevanja.

Če razpoznavalnemu cevovodu na podlagi periokularnega dela s seštevalno fuzijo dodamo razpoznavanje na podlagi beločnice, pride do občutnega izboljšanja (pri verifikacijskih eksperimentih se npr. $VER@0,1FAR$ se iz 0,499 poveča na 0,596, pri identifikaciji pa se Rank-1 zviša iz 0,843 na 0,907). Uspešnost tovrstne fuzije se močno približa uspešnosti razpoznavanja izključno na podlagi šarenice. Med kombinacijami fuzij, ki združujejo po dve modalnosti, se za najbolj uspešno izkaže združevanje šarenice in periokularne regije. Najboljši rezultat pa dobimo ob združevanju vseh treh cevovodov skupaj.

Poglavje 10

Sklepi in nadaljnje delo

Prvi prispevek te naloge je izdelava zbirke SBVPI, za katero bi lahko rekli, da je trenutno najbolj urejena zbirka za raziskave razpoznavanja na podlagi beločnice. Hkrati je edina zbirka, ki je za tovrstne raziskovalne namene v celoti javno dostopna. SBVPI ima v primerjavi s konkurenčnimi zbirkami tudi največ priloženih anotacij, pri čemer imajo posebno vrednost anotacije na nivoju slikovnih elementov, ki smo jim pri izdelavi posvetili veliko časa in pozornosti. V zbirko SBVPI se v nadaljnjem delu lahko dodajo nove identitete, s čimer bi zbirka postala konkurenčna tudi po kriteriju števila identitet oz. slik. Zbirko bi bilo smiselno razširiti tudi s predstavniki različnih etničnih skupin in pridobiti več oseb, ki so stare nad 35 let, saj je večina identitet trenutno stara pod 35 let.

Naslednji prispevek je razvoj cevovoda za razpoznavanje na podlagi vaskularnih struktur v beločnici, ki v celoti temelji na metodah globokega učenja. Celotnega cevovoda za razpoznavanje na podlagi beločnice, ki bi deloval na metodah globokega učenja, v literaturi še nismo zasledili, zato gre najbrž za prvi tovrstni cevovod. V sklopu tega dela so bili razviti algoritmi za segmentacijo beločnice, s katerimi smo na zbirki MASD osvojili prvo mesto na tekmovanjih SSERBC 2017 [20] in SSBC 2018 [44]. Predlagamo tudi postopek za segmentacijo žil, ki dosega vzpodbudne rezultate. V cevovod za razpoznavanje na podlagi beločnice bi bilo v nadaljnjem delu smiselno vključiti še fazo za izboljšanje informacije vaskularnih struktur (angl. vessel enhancement). Pri tem bi kot osnovo lahko vzeli algoritme, ki jih v cevovodih za razpoznavanje na podlagi prstnih odtisov uporabljajo za izboljšanje kvalitete vzorcev. Za razpoznavalno mrežo v delu predlagamo arhitekturo ScleraNET, za katero pokažemo, da deluje bolje kot tradicionalen algoritem SIFT. V nadaljnjem delu bi bilo smiselno narediti tudi primerjavo z drugimi tradicionalnimi algoritmi ter uporabiti drugačne globoke razpoznavalne arhitekture.

Prispevek te naloge je tudi učinkovita večrazredna segmentacija očesa, ki ima potencial pohitrili že obstoječe fuzijske cevovode, ki delujejo na podlagi večih očesnih modalno-

sti. Namesto uporabe večjega števila segmentacijskih modelov, kjer vsak segmentira svoj del očesa, lahko uporabimo le en model, ki je približno enako računsko potraten kot en prvoten segmentacijski model. Na primerih iz zbirke MASD smo pokazali, da večrazredna segmentacija relativno uspešno segmentira v 6 različnih razredov. Za segmentacijo beločnice smo pokazali, da večrazredna segmentacija deluje zelo podobno kakor enorazredna segmentacija beločnice.

V našem delu za vsako modalnost uporabljamo svoj segmentacijski model (beločnico segmentiramo s SegNet modelom, periokularni del z UNet modelom in šarenico z DeepLabV3 modelom). Zaradi pomanjkanja anotacij (katerih pridobitev je izjemno časovno zahtevna) smo med izdelavo tega dela lahko razvijali le segmentacijske algoritme za vsak del očesa posebej. Bolj natančno smo v času izdelave cevovoda za razpoznavanje na podlagi beločnice na zbirki SBVPI označevali le beločnice in izdelali segmentacijski model izključno za ta del, zato da smo lahko izvajali nadaljnje raziskave. Ko smo se lotili naslednjega cevovoda za razpoznavanje na periokularnem delu, smo začeli anotirati le periokularne dele in zgradili smo nov model itd. Vse potrebne anotacije smo tako zbrali šele na koncu. Zato v našem primeru grajenje večrazrednega segmentacijskega algoritma na zbirki SBVPI, ki bi ga uporabili za namene pri fuziji, praktično ni bilo možno zaradi časovnih ovir. Zato bi v nadaljnjem delu lahko (zdaj, ko imamo zbrane vse anotacije) zgradili en večrazreden segmentacijski model, ki bi nadomestil 3 segmentacijske modele, s čimer bi cevovod za končno fuzijo poenostavili.

Naslednji prispevek je primerjava dveh metod globokega učenja za razpoznavanje na podlagi periokularnega dela, kjer primerjamo UNet+ResNet in Deep-PRWIS. V tem poglavju dobimo rezultate, ki se po glavnih zaključkih ujemaajo z delom [16]. Avtorji namreč pokažejo, da model za razpoznavanje na podlagi periokularnega dela lahko implicitno naučimo, da v procesu razpoznavanja ne upošteva informacije iz šarenice oz. beločnice, s čimer se natančnost razpoznavanja izboljša. Hkrati s postopkom Deep-PRWIS poenostavimo razpoznavalni cevovod v fazi testiranja, saj eksplicitna segmentacija ni potrebna.

Za razpoznavanje iz šarenice smo uspešno uporabili že obstoječ cevovod. Kljub temu, da so v zbirki SBVPI zastopane štiri različne smeri pogledov (kar oteži razpoznavanje), smo z uporabo tega cevovoda dobili najvišjo natančnost razpoznavanja (v primerjavi s cevovodi, ki razpoznavanje na podlagi beločnice/periokularne regije). Tak rezultat smo tudi nekako pričakovali, saj je šarenica (kot biometrična modalnost) znana po visokih natančnostih razpoznavanja.

Zadnji prispevek naloge je uporaba vseh treh obravnavanih modalnosti v fuzijskem sistemu na nivoju ocene, kjer smo pokazali, da z združevanjem modalnosti izboljšamo natančnost razpoznavanja, kar velja za oba razpoznavalna scenarija (identifikacijo in verifikacijo). V nadaljnjem delu bi lahko poskusili izvesti fuzijo tudi na drugih nivojih, npr. na nivoju značilk.

Literatura

- [1] G. Kaur and C. K. Verma. Comparative Analysis of Biometric Modalities. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, 4(4):603–613, 2014.
- [2] C. Willoughby, D. Ponzin, S. Ferrari, A. Lobo, K. Landau, and Y. Omid. Anatomy and Physiology of the Human Eye: Effects of Mucopolysaccharidoses Disease on Structure and Function – A Review. *Clinical & Experimental Ophthalmology*, 38:2–11, 2010.
- [3] G. K. Lang. *Ophthalmology: A Pocket Textbook Atlas; revised and enlarged-With 510 illustrations & 50 tables*. Thieme, New York, 2007.
- [4] V. Patil and A. M. Patil. Human Identification Method: Sclera Recognition. *International Journal of Computer Science and Network (IJCSN)*, 6(1):24–29, 2017.
- [5] J. Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 15(11):1148–1161, 1993.
- [6] J. Daugman. How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, 14(1):21–30, 2004.
- [7] J. Daugman. New Methods in Iris Recognition. *IEEE Transactions on Systems, Man, and Cybernetics (TSMC), Part B*, 37(5):1167–1175, 2007.
- [8] D. M. Monro, S. Rakshit, and D. Zhang. DCT-based Iris Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(4):586–595, 2007.
- [9] K. Miyazawa, K. Ito, T. Aoki, K. Kobayashi, and H. Nakajima. An Effective Approach for Iris Recognition Using Phase-based Image Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(10):1741–1756, 2008.
- [10] Z. Sun and T. Tan. Ordinal Measures for Iris Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(12):2211–2226, 2009.

-
- [11] Z. Sun, H. Zhang, T. Tan, and J. Wang. Iris Image Classification Based on Hierarchical Visual Codebook. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(6):1120–1133, 2014.
- [12] I. Nigam, M. Vatsa, and R. Singh. Ocular Biometrics: A Survey of Modalities and Fusion Approaches. *Information Fusion*, 26:1–35, 2015.
- [13] N. D. Kalka, J. Zuo, N. Schmid, and B. Cukic. Image Quality Assessment for Iris Biometric. pages 6202D1–6202D11, 2006.
- [14] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic. Estimating and Fusing Quality Factors for Iris Biometric Images. *IEEE Transactions on Systems, Man, and Cybernetics (TSCM) - Part A: Systems and Humans*, 40(3):509–524, 2010.
- [15] A. Sequeira, L. Chen, J. Ferryman, P. Wild, F. Alonso-Fernandez, J. Bigun, K. Raja, R. Raghavendra, C. Busch, T. Freitas Pereira, et al. Cross-Eyed 2017: Cross-Spectral Iris/Periocular Recognition Database and Competition. In *IEEE International Joint Conference on Biometrics (IJCB)*, 2017.
- [16] H. Proença and J. C. Neves. Deep-PRWIS: Periocular Recognition Without the Iris and Sclera Using Deep Learning Frameworks. *IEEE Transactions on Information Forensics and Security (TIFS)*, 13(4):888–896, 2017.
- [17] Z. Zhao and A. Kumar. Accurate Periocular Recognition Under Less Constrained Environment Using Semantics-assisted Convolutional Neural Network. *IEEE Transactions on Information Forensics and Security (TIFS)*, 12(5):1017–1030, 2017.
- [18] F. Alonso-Fernandez and J. Bigun. Near-Infrared and Visible-Light Periocular Recognition with Gabor Features Using Frequency-Adaptive Automatic Eye Detection. *IET Biometrics*, 4(2):74–89, 2015.
- [19] T. R. Borah, K. K. Sarma, and P. H. Talukdar. Retina Recognition System Using Adaptive Neuro Fuzzy Inference System. In *International Conference on Computers Communications and Control (IC4)*, pages 1–6, 2015.
- [20] A. Das, U. Pal, M. A. Ferrer-Ballester, M. Blumenstein, D. Štepec, P. Rot, Ž. Emeršič, P. Peer, and V. Štruc. SSRBC 2017: Sclera Segmentation and Recognition Benchmarking Competition. *International Joint Conference on Biometrics (IJCB)*, pages 1–6, 2017.
- [21] S. Alkassar, W. L. Woo, S. Dlay, and J. Chambers. Sclera Recognition: On the Quality Measure and Segmentation of Degraded Images Captured Under Relaxed Imaging Conditions. *IET Biometrics*, 6(4):266–275, 2016.

-
- [22] S. Alkassar, W. L. Woo, S. S. Dlay, and J. A. Chambers. Robust Sclera Recognition System with Novel Sclera Segmentation and Validation Techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems (TSMC)*, 47(3):474–486, 2017.
- [23] F. Sadikoglu and S. Uzelaltinbulat. Biometric Retina Identification Based on Neural Network. *12th International Conference on Application of Fuzzy Systems and Soft Computing (ICAFS)*, 102:26–33, 2016.
- [24] W. Goebel and T. Kretzchmar-Gross. Retinal Thickness in Diabetic Retinopathy: A Study Using Optical Coherence Tomography (OCT). *Retina*, 22(6):759–767, 2002.
- [25] T. de Freitas Pereira and S. Marcel. Periocular Biometrics in Mobile Environment. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, 2015.
- [26] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro. Fusing Iris and Periocular Information for Cross-sensor Recognition. *Pattern Recognition Letters*, 57:52–59, 2015.
- [27] S. Bakshi, P. K. Sa, H. Wang, S. S. Barpanda, and B. Majhi. Fast Periocular Authentication in Handheld Devices with Reduced Phase Intensive Local Pattern. *Multimedia Tools and Applications*, pages 1–29, 2017.
- [28] N. Aginako, M. Castrillón-Santana, J. Lorenzo-Navarro, J. M. Martínez-Otzeta, and B. Sierra. Periocular and Iris Local Descriptors for Identity Verification in Mobile Applications. *Pattern Recognition Letters*, 91:52–59, 2017.
- [29] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [30] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556, 2014.
- [31] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet: A Deep Convolutional Encoder–Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 39(12):2481–2495, 2017.
- [32] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha. Deep Representations for Iris, Face, and Fingerprint Spoofing Detection. *IEEE Transactions on Information Forensics and Security (TIFS)*, 10(4):864–879, 2015.

- [33] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan. Iris Recognition with Off-the-Shelf CNN Features: A Deep Learning Perspective. *IEEE Access*, 6:18848–18855, 2018.
- [34] N. Liu, M. Zhang, H. Li, Z. Sun, and T. Tan. DeepIris: Learning Pairwise Filter Bank for Heterogeneous Iris Verification. *Pattern Recognition Letters*, 82:154–161, 2016.
- [35] A. Gangwar and A. Joshi. DeepIrisNet: Deep Iris Representation with Applications in Iris Recognition and Cross-Sensor Iris Recognition. *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2301–2305, 2016.
- [36] A. Das, U. Pal, M. Blumenstein, and M. A. Ferrer-Ballester. Sclera Recognition – A Survey. In *Asian Conference on Pattern Recognition (ACPR)*, pages 917–921, 2013.
- [37] P. Radu, J. Ferryman, and P. Wild. A Robust Sclera Segmentation Algorithm. *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1–6, 2015.
- [38] Z. Zhou, E. Y. Du, N. L. Thomas, and E. J Delp. A New Human Identification Method: Sclera Recognition. *IEEE Transactions on Systems, Man, and Cybernetics (TSMC) -Part A: Systems and Humans*, 42(3):571–583, 2012.
- [39] V. Gottemukkula, S. K. Saripalle, S. P. Tankasala, R. Derakhshani, R. Pasula, and A. Ross. Fusing Iris and Conjunctival Vasculature: Ocular Biometrics in the Visible Spectrum. *IEEE Conference on Technologies for Homeland Security (HST)*, pages 150–155, 2012.
- [40] R. Derakhshani, A. Ross, and S. Crihalmeanu. A New Biometric Modality Based on Conjunctival Vasculature. *Artificial Neural Networks in Engineering (ANNIE)*, pages 1–8, 2006.
- [41] R. Derakhshani and A. Ross. A Texture-based Neural Network Classifier for Biometric Identification Using Ocular Surface Vasculature. *International Joint Conference on Neural Networks (IJCNN)*, pages 2982–2987, 2007.
- [42] A. Das, U. Pal, M. A. Ferrer-Ballester, and M. Blumenstein. SSBC 2015: Sclera Segmentation Benchmarking Competition. *IEEE 7th International Conference on Biometrics Theory (BTAS), Applications and Systems*, pages 1–6, 2015.
- [43] A. Das, U. Pal, M. A. Ferrer-Ballester, and M. Blumenstein. SSRBC 2016: Sclera Segmentation and Recognition Benchmarking Competition. *International Conference on Biometrics (ICB)*, pages 1–6, 2016.

- [44] A. Das, U. Pal, M. A. Ferrer-Ballester, M. Blumenstein, D. Štepec, P. Rot, P. Peer, and V. Štruc. SSBC 2018: Sclera Segmentation Benchmarking Competition. *International Conference on Biometrics (ICB)*, pages 303–308, 2018.
- [45] R. Mesbah, B. McCane, and S. Mills. Conditional Random Fields Incorporate Convolutional Neural Networks for Human Eye Sclera Semantic Segmentation. *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 768–773, 2017.
- [46] D. R. Lucio, R. Laroca, E. Severo, A. S. Britto Jr, and D. Menotti. Fully Convolutional Networks and Generative Adversarial Networks Applied to Sclera Segmentation. *CoRR*, vol. *abs/1806.08722*, 2018.
- [47] H. Proença and L. Alexandre. UBIRIS: A Noisy Iris Image Database. In *International Conference on Image Analysis and Processing (ICIAP)*, pages 970–977. Springer, 2005.
- [48] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A Alexandre. The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(8):1529–1535, 2010.
- [49] M. S Hosseini, B. N Araabi, and H. Soltanian-Zadeh. Pigment Melanin: Pattern for Iris Recognition. *IEEE Transactions on Instrumentation and Measurement (TIM)*, 59(4):792–804, 2010.
- [50] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile Iris Challenge Evaluation (MICHE)-I, Biometric Iris Dataset and Protocols. *Pattern Recognition Letters*, 57:17–23, 2015.
- [51] A. Das, U. Pal, M. A. Ferrer-Ballester, and M. Blumenstein. A New Method for Sclera Vessel Recognition Using OLBP. In *Biometric Recognition*, pages 370–377. Springer, 2013.
- [52] E. G. Maxwell and C. Tripti. A Comparison Between Contrast Limited Adaptive Histogram Equalization and Gabor Filter Sclera Blood Vessel Enhancement Techniques. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(4):22–5, 2013.
- [53] S. P. Tankasala, P. Doynov, R. R. Derakhshani, A. Ross, and S. Crihalmeanu. Biometric Recognition of Conjunctival Vasculature Using GLCM Features. In *2011 International Conference on Image Information Processing (ICIIP)*, pages 1–6, 2011.
- [54] A. Das, U. Pal, M. A. Ferrer-Ballester, and M. Blumenstein. Sclera Recognition Using Dense-SIFT. In *2013 13th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 74–79. IEEE, 2013.

- [55] A. Das, U. Pal, M. A. Ferrer-Ballester, and M. Blumenstein. Fuzzy Logic Based Sclera Recognition. In *2014 IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 561–568, 2014.
- [56] U. Park, A. Ross, and A. K. Jain. Periocular Biometrics in the Visible Spectrum: A Feasibility Study. In *2009 IEEE 3rd International Conference on Biometrics (ICB): Theory, Applications, and Systems*, pages 1–6, 2009.
- [57] C. N. Padole and H. Proença. Periocular Recognition: Analysis of Performance Degradation Factors. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 439–445, 2012.
- [58] F. Alonso-Fernandez and J. Bigun. A Survey on Periocular Biometrics Research. *Pattern Recognition Letters*, 82:92–105, 2016.
- [59] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh. Periocular Biometrics: When Iris Recognition Fails. In *International Conference on Biometrics Theory (BTAS)*, pages 1–6, 2010.
- [60] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 947–954, 2005.
- [61] P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(10):1090–1104, 2000.
- [62] CASIA Iris Image Database. <http://biometrics.idealtest.org/>.
- [63] UPOL Iris Image Database. <http://phoenix.inf.upol.cz/iris/>.
- [64] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On Cross Spectral Periocular Recognition. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 5007–5011. IEEE, 2014.
- [65] U. Park, R. R. Jillela, A. Ross, and A. K. Jain. Periocular Biometrics in the Visible Spectrum. *IEEE Transactions on Information Forensics and Security (TIFS)*, 6(1):96–106, 2011.
- [66] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard. Soft Biometric Classification Using Periocular Region Features. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–7, 2010.

- [67] Z. Ali, U. Park, J. Nang, J. S. Park, T. Hong, and S. Park. Periocular Recognition Using uMLBP and Attribute Features. *KSII Transactions on Internet and Information Systems (TIIS)*, 11(12):6133–6151, 2017.
- [68] S. Patil, S. Gudasalamani, and N. C. Iyer. A Survey on iris Recognition System. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pages 2207–2210, 2016.
- [69] M. Rifaee, M. Abdallah, and B. Okosh. A Short Survey for Iris Images Databases. *International Journal of Multimedia and its Applications*, 8(2):1–14, 2017.
- [70] V. G. Garagad and N. C. Iyer. A Novel Technique of Iris Identification for Biometric Systems. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 973–978, 2014.
- [71] H. Rai and A. Yadav. Iris Recognition Using Combined Support Vector Machine and hamming Distance Approach. *Expert Systems with Applications*, 41(2):588–593, 2014.
- [72] N. U. Ahmed, S. Cvetkovic, E. H. Siddiqi, A. Nikiforov, and I. Nikiforov. Using Fusion of Iris Code and Periocular Biometric for Matching Visible Spectrum Iris Images Captured by Smart Phone Cameras. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 176–180, 2016.
- [73] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender Classification from the Same Iris Code Used for Recognition. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(8):1760–1770, 2016.
- [74] S. Minaee, A. A. Abdolrashidi, and Y. Wang. An Experimental Study of Deep Convolutional Features for Iris Recognition. *CoRR*, abs/1702.01334, 2017.
- [75] M. Vatsa, R. Singh, A. Ross, and A. Noore. Quality-Based Fusion for Multichannel Iris Recognition. In *2010 20th International Conference on Pattern Recognition (ICPR)*, pages 1314–1317, 2010.
- [76] A. Kumar and A. Passi. Comparison and Combination of Iris Matchers for Reliable Personal Authentication. *Pattern Recognition*, 43(3):1016–1026, 2010.
- [77] C. W. Tan and A. Kumar. Human Identification from at-a-distance Images by Simultaneously Exploiting Iris and Periocular Features. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 553–556, 2012.
- [78] A. K. Jain, J. Mao, and K. M. Mohiuddin. Artificial Neural Networks: A Tutorial. *Computer*, 29(3):31–44, 1996.

- [79] B. Blaus. Figure showing Neuron. Creative Commons 3.0 (<https://creativecommons.org/licenses/by/3.0/>). Dostopano preko <https://en.wikipedia.org/wiki/Neuron/>, 2013.
- [80] W.S. McCulloch and W. Pitts. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Mathematical Biophysics*, 5(1):115–133, 1943.
- [81] A. Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 1 edition, 2017.
- [82] R. Rosenblatt. *Principles of Neurodynamics*. Spartan Books, New York, 1 edition, 1962.
- [83] X. Glorot, A. Bordes, and Y. Bengio. Deep Sparse Rectifier Neural Networks. In G. Gordon, D. Dunson, and M. Dudik, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15, pages 315–323, 2011.
- [84] A. Ng. Stanford University and Coursera Course: Machine Learning. [online] [coursera.org](https://www.coursera.org/learn/machine-learning) Dostopno na: <https://www.coursera.org/learn/machine-learning>, 2014.
- [85] D. Rumelhart, G. Hinton, and R. Williams. Learning Representations by Back-propagating Errors. *Nature*, 323(6088):533–540, 1986.
- [86] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [87] D. Hubel and T. Wiesel. Receptive Fields of Single Neurones in the Cat’s Striate Cortex. *The Journal of Physiology*, 148(3):574–591, 1959.
- [88] A. Karpathy. CS231n Convolutional Neural Networks for Visual Recognition. [online] [CS231n.github.io](http://cs231n.github.io). Dostopno na: <http://cs231n.github.io>, 2017.
- [89] P. Veličković. Deep Learning for Complete Beginners: Convolutional Neural Networks with Keras. [online] cambridgespark.com Dostopno na: <https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html>, 2017.
- [90] R. Pinto. Figure: obtaining different feature maps using multiple filters. [online] [stackoverflow.com](http://user-image.logdown.io/user/13673/blog/12890/post/302641/V21DnAAeTKiOirZRFKhT_dl2.png) Dostopno na: http://user-image.logdown.io/user/13673/blog/12890/post/302641/V21DnAAeTKiOirZRFKhT_dl2.png, 2016.

- [91] M. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision (ECCV)*, pages 818–833. Springer, Cham, 2014.
- [92] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. F. Li. ImageNet Large Scale Visual Recognition Challenge. *CoRR*, abs/1409.0575, 2014.
- [93] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [94] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015.
- [95] G. Lin, A. Milan, C. Shen, and I. D. Reid. RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5168–5177, 2017.
- [96] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. *Medical Image Computing and Computer-Assisted Intervention (MICCAI): 18th International Conference, Proceedings, Part III*, pages 234–241. Springer International Publishing, Cham, 2015.
- [97] P. Rot, V. Štruc, Ž. Emeršič, and P. Peer. Deep Multi-class Eye Segmentation for Ocular Biometrics. In *IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, 2018.
- [98] X. Yin and X. Liu. Multi-task Convolutional Neural Network for Pose-invariant Face Recognition. *IEEE Transactions on Image Processing (TIP)*, 2017.
- [99] J. Lozej, B. Meden, V. Štruc, and P. Peer. End-to-End Iris Segmentation Using U-Net. In *IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, 2018.
- [100] Ž. Emeršič, L. Gabriel, V. Štruc, and P. Peer. Convolutional Encoder–Decoder Networks for Pixel-wise Ear Detection and Segmentation. *IET Biometrics*, 7(3):175–184, 2018.
- [101] Ž. Emeršič, D. Štepec, V. Štruc, and P. Peer. Training Convolutional Neural Networks with Limited Training Data for Ear Recognition in the Wild. In *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2017.

- [102] J. Križaj, V. Štruc, and S. Dobrišek. Combining 3d Face Representations Using Region Covariance Descriptors and Statistical Models. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–7. IEEE, 2013.
- [103] V. Štruc. The PhD Face Recognition Toolbox: Toolbox Description and User Manual. *Faculty of Electrical Engineering Ljubljana*, 2012.
- [104] B. Vesnicer, J. Žganec Gros, N. Pavešić, and V. Štruc. Face Recognition Using Simplified Probabilistic Linear Discriminant Analysis. *International Journal of Advanced Robotic Systems (IJARS)*, 9(5):180–190, 2012.
- [105] Ž. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. E Boulton, R. Safdaii, Y. Zhou, et al. The Unconstrained Ear Recognition Challenge. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 715–724. IEEE, 2017.
- [106] X. Glorot and Y. Bengio. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics (AISTATS)*, pages 249–256, 2010.
- [107] V. Štruc and N. Pavešić. Phase Congruency Features for Palm-print Verification. *IET Signal Processing*, 3(4):258–268, 2009.
- [108] T. Savič and N. Pavešić. Personal Recognition Based on an Image of the Palmar Surface of the Hand. *Pattern Recognition*, 40(11):3152–3163, 2007.
- [109] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. Residual Attention Network for Image Classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6458, 2017.
- [110] D. Riccio, N. Brancati, M. Frucci, and D. Gagnaniello. An Unsupervised Approach for Eye Sclera Segmentation. In *Iberoamerican Congress on Pattern Recognition (CIARP)*, pages 550–557. Springer, 2017.
- [111] R. Gonzalez and R. Woods. Image Processing. *Digital image processing*, 2, 2007.
- [112] D. G. Lowe. Object Recognition from Local Scale-invariant Features. In *International Journal of Computer Vision (ICCV)*, volume 2, pages 1150–1157. Ieee, 1999.
- [113] D. R. Ibrahim, A. A. Tamimi, and A. M. Abdalla. Performance Analysis of Biometric Recognition Modalities. In *2017 8th International Conference on Information Technology (ICIT)*, pages 980–984, 2017.
- [114] J. Lozej. Iris Recognition Using Deep Learning. Master’s thesis, Faculty of Computer and Information Science, University of Ljubljana, Slovenia, 2018.

- [115] A. Jain, A. Ross, and K. Nandakumar. *Introduction to Biometrics*. Springer US, 2011.