

PRIMERJAVA ČLOVEŠKE IN RAČUNALNIŠKE PREPOZNAVE NA OSNOVI GLOBOKEGA UČENJA

Raziskovalno področje: RAČUNALNIŠTVO IN TELEKOMUNIKACIJE

Raziskovalna naloga

Primož Perko, Mateo Paulec, Miha Govedič

Mentor: Mitja Osojnik

II. gimnazija Maribor

Maribor, februar 2020

54. srečanje mladih raziskovalcev Slovenije 2020

PRIMERJAVA ČLOVEŠKE IN RAČUNALNIŠKE PREPOZNAVE NA OSNOVI GLOBOKEGA UČENJA

Raziskovalno področje: RAČUNALNIŠTVO IN TELEKOMUNIKACIJE

Raziskovalna naloga

Primož Perko, Mateo Paulec, Miha Govedič

Mentor: Mitja Osojnik

II. gimnazija Maribor

Maribor, 2020

KAZALO VSEBINE

POVZETEK	6
1. UVOD	10
1.1 Namen	10
1.2 Hipoteze	11
2 TEORIJA	12
2.1 Človeška prepoznavna	12
2.1.1 Biološka	12
2.1.2 Psihološka (kognitivna)	13
2.2 Računalniška prepoznavna	15
2.2.1 Modeli globokega učenja	16
3 ORODJA ZA PREPOZNAVNO	17
3.1 Knjižnica	17
3.2 Model	17
3.3 Baza podatkov	19
4 METODOLOGIJA DELA	20
4.1 Izbor slik	20
4.2 Človeška metoda	20
4.3 Računalniška metoda	21
5 REZULTATI	22
5.1 Slike 1, 2, 3	23
5.2 Slika 4	24
5.3 Slika 5	25
5.4 Slika 6	26
5.5 Slika 7	26
5.6 Slika 8	26
5.7 Sliki 9 in 10	29
5.8 Slika 11	29
5.9 Slika 12	29
5.10 Slika 13	32
5.11 Slika 14	33
5.12 Slika 15	33
6 ZAKLJUČKI	35

6.1	Kategorija 1	35
6.2	Kategorija 2	36
6.3	Kategorija 3	38
7	ZAKLJUČEK	40
8	DRUŽBENA ODGOVORNOST	42
9	PRILOGE	43
9.1	Priloga 1 – slike, sodelujoče v raziskavi	43
9.2	Priloga 2 – rezultati računalniške primerjave	48
9.2.1	Slike 1/12	48
9.3	Priloga 3 – tabela rezultatov intervjuja	57
	Bibliografija	58

KAZALO GRAFIKONOV

GRAFIKON 1: SLIKE 1, 2, 3	23
GRAFIKON 2: SLIKE 4, 5, 6	25
GRAFIKON 3: SLIKI 7, 8	28
GRAFIKON 4: SLIKI 9, 10	29
GRAFIKON 5: SLIKI 11, 12	31
GRAFIKON 6: SLIKA 13, 14, 15	34

KAZALO SLIK

SLIKA 1: PRIKAZ PRENOSA SLIKE V MOŽGANE (TYLEY, 2015)	12
SLIKA 2: OSNOVNA DELITEV STROJNEGA UČENJA (PANT, 2019)	15
SLIKA 3: PRIKAZ UPORABNOSTI RAZLIČNIH UČNIH ALGORITMOV (ELFOULY, 2019)	15
SLIKA 4: ILUSTRACIJA NEVRONSKE MREŽE (SAMUEL, 2019)	17
SLIKA 5: SHEMA DELOVANJA SSD ALGORITMA (SOROKINA, 2017)	18
SLIKA 6: PRIKAZ KATEGORIJE BAZE PODATKOV COCO (COCO, 2019)	19
SLIKA 7: RAČUNALNIKOV REZULTAT SLIKE 12 (OCAMPO, 2020)	31

KAZALO PREGLEDNIC

PREGLEDNICA 1: SLIKA 1, 2, 3	22
PREGLEDNICA 2: SLIKE 4, 5, 6	24
PREGLEDNICA 3: SLIKI 7, 8	26
PREGLEDNICA 4: SLIKI 7, 8	26
PREGLEDNICA 5: SLIKE 9, 10, 11, 12	28
PREGLEDNICA 6: SLIKE 13, 14, 15	32
PREGLEDNICA 7: SLIKE 13, 14, 15	33

POVZETEK

Z napredkom računalniških tehnologij, se odpirajo nove možnosti, s katerimi se srečujemo v vsakdanjem življenju. Ob vse večji prisotnosti umetne inteligence nas je zanimalo, kako dobro lahko računalnik prepozna slike in kako se ob tem primerja s človekom. Pripravili smo reprezentativni vzorec slik in za klasifikacijo, z metodo strukturiranega intervjuja, prosili ljudi. Iste slike smo klasificirali tudi s pomočjo algoritma, ki bazira na nevronske mrežah in je bil naučen z metodo globokega učenja. Kot smo predpostavili na začetku, je računalnik pravilno razbral predmete, ni pa mogel razložiti konteksta. Prav tako računalnik ni razložil slik, na katerih iskane vsebine niso bile povsem očitne. Ugotovili smo, da računalnik sicer dobro prepoznavna predmete, ni pa zmožen tvoriti povezav med posameznimi zaznavami iz katerih bi bilo mogoče razbrati kontekst. Poglabljena zaznava je torej še vedno izključno domena ljudi, saj intervjuvanci z nalogami niso imeli težav.

ZAHVALA

Radi bi se zahvalili mentorju, ki nas je na dolgi poti raziskovanja spremljal in usmerjal.

Pomagal nam je pri izdelavi naloge, vse od oblikovanja ideje pa do zaključne zgradbe.

Zavaliti se moramo tudi staršem ter bratom, ki so nam nudili podporo in pomoč, da smo lahko dokončali raziskovalno delo.

Posebna zahvala gre tudi vsem, ki so si vzeli čas za naš intervju, saj smo le tako lahko postavili temelje naloge.

Zadnja zahvala gre tebi, bralec, ki si si vzel čas, da si pogledal v vsebino te naloge in pregledal naše delo.

1. UVOD

Računalniki. Orodja, ki jim človeštvo posveča toliko pozornosti in ki upravljajo naše vsakodnevno življenje imajo zdaj sposobnost, da razmišljajo kot človek. Možno je namreč vzpostaviti nevronske mreže, ki delujejo podobno kot človeški možgani, zato so računalniki kot še nikoli doslej primerljivi s človeškim razmišljanjem.

1.1 Namen

Namen te raziskovalne naloge je raziskati, kako so računalniki primerljivi z nami na enem od področij, ki ga ljudje uporabljamo vsak dan, to je pri prepoznavi objektov. Zanima nas, ali se računalnik lahko z natančnostjo kosa s človeškim umom, ne le v prepoznavi preprostih objektov na sliki, marveč tudi v bolj zapletenih odnosih med objekti znotraj ene slike. Raziskati želimo tudi, kako se obnese računalnik pri interpretaciji slik, ki človeško prepoznavo potisnejo do roba zmognosti. Zato smo se odločili primerjavo opraviti tudi na slikah, za katere ljudje trdijo, da lahko na njih vidijo več kot en objekt; torej na slikah z optičnimi iluzijami, ki vsebujeta dve sliki v eni.

Celotna raziskovalna naloga temelji na načelu primerjave. Primerjali smo namreč rezultate, ki so jih dobili človeški osebki, in rezultate, ki jih je dobil računalnik. Za nalogo je bilo primerjanje pomembno, saj se računalniki z nevronskimi mrežami in *algoritmi globokega učenja (deep learning)* skušajo približati človeškemu delovanju. Ker je vid in prepoznavna ena pomembnejših človeških lastnosti, je primerjava na področju, ki ga raziskujemo, še posebej pomembna. Neposredna primerjava nam ponuja edinstven vpogled v razvoj tehnologije in hkrati skuša osmisliti koncept človeške prepoznavne objektov. Omogoča nam, da ugotovimo, kako blizu človeški prepoznavi so danes dostopne tehnologije, in tako pomaga ugotoviti, ali so lahko uporabne v praktičnem okolju.

1.2 Hipoteze

Za raziskavo smo si zastavili tri glavne hipoteze, vsaka ustreza svoji vrsti slik(e) in napoveduje, kako se bo obnesel računalnik v eni izmed treh kategorij, med katere smo uvrstili vsako sliko v eksperimentu. Testirali smo na računalniku in skupini ljudi.

Postavili smo naslednje hipoteze:

- H1: Računalnik bo slike z enim objektom označil z enako oznako kot človek.
- H2: Na slikah z vsebino, katere elementi in objekti imajo med sabo dinamične odnose (na primer »*hranjenje psa*«), bo računalnik znal opredeliti le posamezne objekte znotraj slike, človeški opazovalec pa bo znal prepoznati dogajanje in ga na preprost način opisati.
- H3: Na slikah, v katerih je moč prepoznati dva ali več objektov, bo odstotek ljudi, ki na sliki prepozna enega od objektov, približno enak odstotku prepričanja računalnika, da vidi na sliki ta isti objekt. (+/- 5 %).

Hipoteza 1 pomeni, da se lahko računalniki v okolju z relativno malo motnjami in izrazitim slikovnim materialom obnesejo enako dobro kot ljudje. Hipoteza 2 pomeni, da se računalnik v kompleksnih odnosih med objekti ne znajde in lahko le na osnovnem nivoju prepozna zaključene elemente in objekte. Hipoteza 3 pomeni, da lahko računalnik s približno enako gotovostjo kot človek prepozna slike, ob katerih se človeška prepozna zmede.

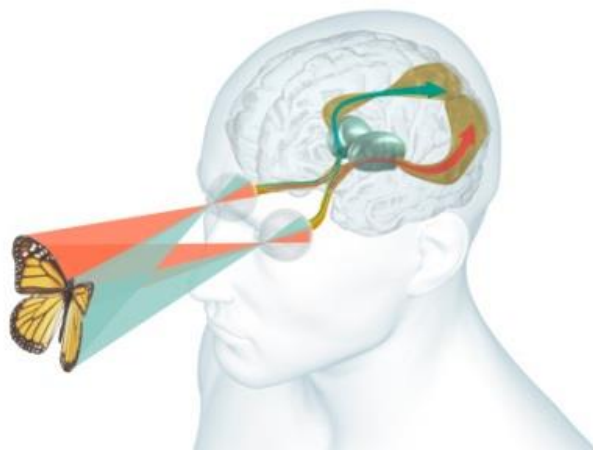
2 TEORIJA

Za primerjavo je bilo potrebno zbrati dva vzorca podatkov: računalniškega in človeškega. Ker računalnik uporablja nevronske mreže in algoritem, ki izpiše »odstotek prepričanosti« za dano sliko, je lahko ta del opravil le en računalnik.

2.1 Človeška prepoznavna

2.1.1 Biološka

Človek za zaznavo objektov v naravi uporablja oko in možgane. Oči so se skozi 100 milijonov let razvijale vse od temno-svetlega senzorja do zgradbe, ki jo imamo danes. Struktura in delovanje očesa je podobno fotoaparatu. Ima odprtino, skozi katero pride svetloba, lečo, ki izostri sliko in mrežnico, ki ustvari sliko. Količino svetlobe, ki



Slika 1: Prikaz prenosa slike v možgane (Tyley, 2015)

vstopa v oko, nadzoruje šarenica, ki z gladko mišico spreminja velikost zenice. Le-ta se širi pri šibki svetlobi in oži pri močni. Svetloba gre najprej skozi roženico in nato v lečo. Roženica in leča skupaj lomita svetlobo in jo centrirata v točko na mrežnici na koncu očesa (rumena pega), ki vsebuje največ receptorjev, in ima zato taka slika najboljšo ločljivost. Ta točka je zelo majhna (velikost nohta na palcu ob iztegnjeni roki), zato je naše uporabna ločljivost ob danem trenutku zelo majhna. To pa nadomestimo z neprestanim gibanjem očesa. Mrežnica je prekrita z receptorji, občutljivimi na svetlobo. Vsak izmed receptorjev vsebuje molekule pigmenta, ki se spremenijo ob stiku s svetlobo, pri tem pošljejo sporočilo možganom preko optičnih živcev, kot je razvidno na sliki 1. Signali gredo najprej skozi internevrone in nevrone, imenovane ganglionske celice. Nevroni filtrirajo nekatere informacije, ki so nepotrebne, predno dosežejo končni cilj. Možganom predajo informacije optični živci. Dva optična živca se križata in srečata v tako imenovani „optic chiasm“. Tukaj se signali z levega dela obeh oči pretvorijo na levo stran možganov in obratno. To omogoči primerjavo in združevanje obeh slik z obeh oči. Signali vstopijo v

možgane skozi talamus, ki loči informacije na dva dela. En del vsebuje barvo in detajle, drugi pa premikanje in kontrast. Sporočilo se nato premakne na zadnji del možganov v vizualni korteks. Res je, da so oči zmožne projicirati samo dvodimenzionalne slike, ampak ker so naše oči približno 5 centimetrov narazen, je vsaka slika opazovana iz malo drugačnega zornega kota. S pomočjo naših možganov, ki ti dve sliki primerjajo, lahko vidimo globino. Vsako oko ima okoli 6 do 7 milijonov stožčastih celic, pri čemer ima vsak enega izmed treh proteinov, imenovanih opsin. Kot že omenjeno, opsin spremeni obliko v stiku s svetlobo in pošlje signal. Polovica celic se odziva na rdečo svetlobo, tretjina na zeleno in samo dva odstotka na modro, zato je naš vid izostren okoli rumeno zelenega spektra. Na slikah, uporabljenih za raziskovalno nalogo, so ljudje opazili različne vsakodneвне objekte, ki so jih lahko povezali z imeni, ki so se jih naučili. Objekt, ki so ga poimenovali, je bil odvisen od tega, na kateri del ekrana so najprej pogledali in njegove osvetljenosti. (Tyley, 2015)

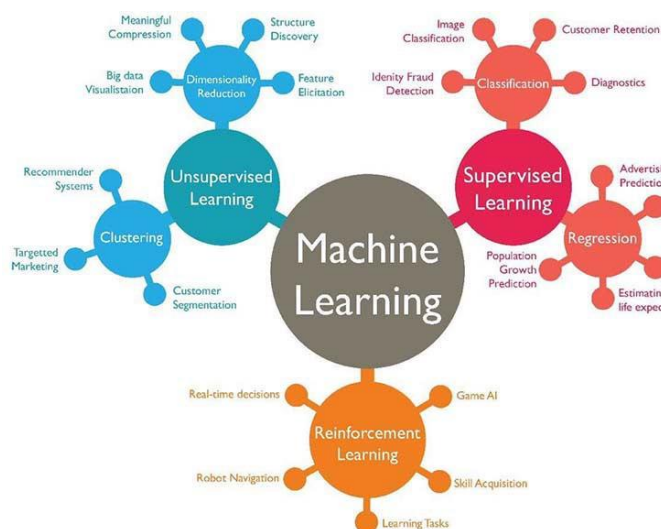
2.1.2 Psihološka (kognitivna)

Človeška sposobnost prepoznavanja je vzbujala zanimanje med znanstveniki že mnoga leta. Nazadnje so ugotovili, da se naše prepoznavanje objektov deli na dva, ne neposredno povezana segmenta. Spomin oziroma »recollection«, ki je dokaj počasen proces in nam prinaša kvalitativne podatke, ki smo se jih naučili, na primer: kdaj in kje smo se kaj učili. V nasprotju s spominom je seznanjenost oziroma »familiarity«, ki temelji na kvantitativnih podatkih, kolikokrat smo že kaj slišali in podobno. V začetku so znanstveniki mislili, da sta ta dva procesa med seboj povezana, a so ugotovili, da delujeta vsak posebej in pri nekaterih prepoznavanjih uporabljamo samo en ali drug segment, in ne obeh hkrati. Spomin pridobi več koristi iz elaborativnega kodiranja, kot seznanjenost, ampak za njegovo nadgradnjo porabi več časa in pozornosti.^[TM1] Pri večini primerov pa naši možgani uporabljajo oba procesa, saj je s tem omogočena najboljša prepoznavnost oziroma izražanje. Na primeru, kjer vidimo človeka, ki se nam zdi znan, ampak se ne spomnimo, od kod ali kako mu je ime. Vidimo, da morata za najboljšo prepoznavnost biti usklajena oziroma v pogonu tako spomin kot seznanjenost. V nalogi bo v primeru kategorije 1 najbolj delovala seznanjenost, medtem ko bosta pri kategoriji 2 imeli vpliv tako seznanjenost kot spomin. (Rugg, 2003)

Informacije o zaznavi slik je najlažje pridobiti s strukturiranim intervjujem.

»Pri strukturiranem intervjuju je proces pridobivanja podatkov poenoten pri vseh. Vsi intervjuvanci odgovarjajo na enaka vprašanja, pri vseh intervjujih so enaka navodila, vedno uporabimo enak postopek zapisovanja odgovorov, pri vseh intervjujih je enak odnos med tistim, ki intervjuja, in intervjuvancem. Tisti, ki vodijo intervju, naj bi bili povsem nevtralni, iz njihovih odgovorov naj ne bi bilo mogoče razbrati osebnega stališča do proučevane teme, če intervjuje izvaja več oseb, morajo biti vsi za to delo enako primerni, izenačiti je potrebno čas in prostor izvajanja intervjuja.« (Agencija za raziskovanje RS, 2009)

2.2 Računalniška prepoznavna



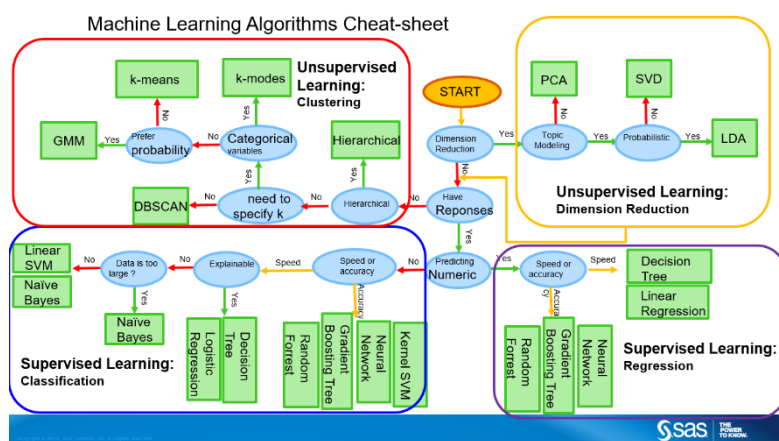
celotna stvar osnovana na nevronskih mrežah. Te delujejo s povezovanjem nevronov. Povezave se okrepijo, če podobni

obrodijo podobne rezultate.

»Algoritmi strojnega učenja (machine learning) dopuščajo računalnikom, da se lahko učijo iz podanih podatkov in se celo izboljšajo, ne da so za to eksplicitno programirani.«(Samuel, 2019)

»Strojno učenje je kategorija algoritma, ki omogoča aplikacijam programske opreme, da so lahko bolj natančne pri predvidevanju izidov, ne da so za to eksplicitno programirane. Osnova strojnega učenja je grajenje algoritmov, ki lahko sprejmejo podatke in uporabijo statistično analizo, da predvidijo končni rezultat in se sproti učijo, ko dostopajo do novih podatkov.« (Pant, 2019)

Globoko učenje deluje podobno kot možgani, saj pri tem tipu učenja uporabljamo veliko število nevronskih točk, ki jih povezujemo v nevronska omrežja. Možgani delujejo na takšen način, saj se nevroni povezujejo z drugimi nevroni in so tako sposobni kompleksnih kalkulacij.



Slika 3: Prikaz uporabnosti različnih učnih algoritmov (Elfouly, 2019)

2.2.1 Modeli globokega učenja

Da lahko računalnik z globokim učenjem pravilno poišče in klasificira objekt, potrebuje ustrezen model. Model dopušča algoritmu, da iz osnovnih podatkov (slike) pride do končne zaznave. Za nalogo prepoznavne in iskanja objektov v sliki poznamo tri pomembne družine modelov.

R-CNN (*region proposed-convolutional neural network*) deluje tako, da najprej v sliki poišče dele, ki bi lahko predstavljali objekt, in jih loči od ozadja. Nato vsakega izmed delov primerja z vsako možno kategorijo, ki jo ima shranjeno v podatkovni bazi, in na načelih statistične analize predvidi, katera izmed kategorij najbolj ustreza objektu. (Elfouly, 2019)

YOLO (*you only look once*) je metoda zaznave objektov, ki sliko le enkrat pogleda in določi objekte v njej, tako da istočasno zazna in označi vse možne okvirčke, ki bi lahko bili objekt. Izdela tudi »mrežo verjetnosti«, ki je $N \times M$ velika mreža, ki sliko razdeli v več delov, v vsakem pa poišče kategorijo, kateri vsebina dela najbolj ustreza. Ta dva podatka pa potem primerja in smiselno poveže ter poišče korelacijo, in tako identificira objekte na sliki. (Redmon, 2016)

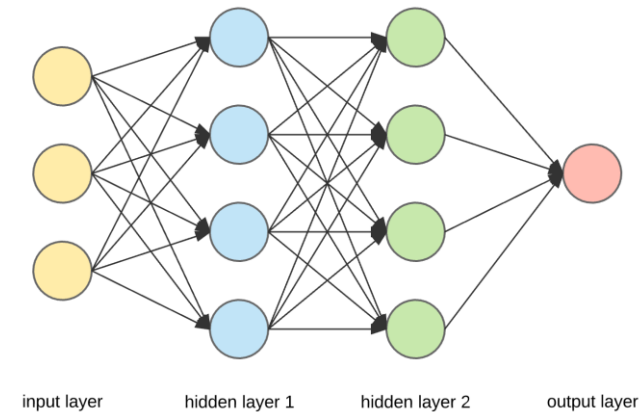
SSD (*single shot detection*) je družina modelov, iz katere izhaja tudi model, ki smo ga uporabili mi. Izmed naštetih je najbolj napredna in tudi najsodobnejša. Na področju se najbolj približa človeški zaznavi, kar je razlog, da smo jo izbrali za raziskavo. Vredno je omeniti, da tudi YOLO spada pod oznako *single shot detector*, a deluje malo drugače kot naš model.

(Khandelwal, Towards data science, 2019)

3 ORODJA ZA PREPOZNAVO

Za našo raziskavo smo uporabili kombinacijo več delov programske opreme, ki skupaj tvorijo celoto, ki lahko prepozna slike.

3.1 Knjižnica



Slika 4: Ilustracija nevronske mreže (Samuel, 2019)

Najprej smo potrebovali programsko knjižnico z vsemi potrebnimi funkcijami in operacijami za strojno učenje (*tf*, *tf.keras* ...). Izbrali smo knjižnico Tensor Flow, ki ga je za Googlovo uporabo razvila ekipa Google Brain in je odprta za prosto dostopno uporabo od leta 2015. Knjižnica podpira programski jezik Python 3.7, s katerim smo delali. (Google, 2019)

S Tensorflow pa smo uporabili tudi knjižnico Keras, ki poskrbi za ustvarjanje nevronske mreže, ki so bile trenirane pri izdelavi modela. Keras dopušča izdelavo zapletenih nevronske mreže (convolutional neural networks), ki so prav tako prosto dostopne. (Keras, 2019)

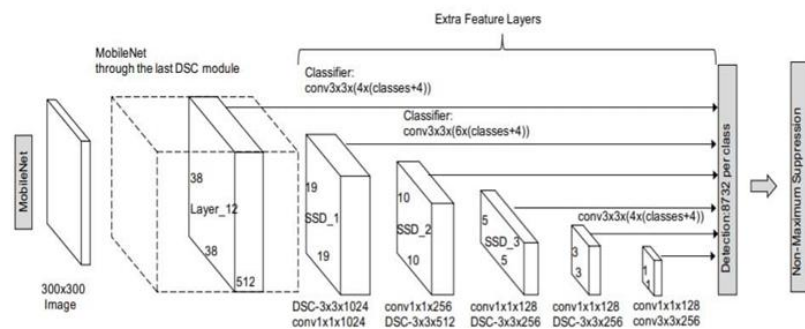
3.2 Model

Potrebovali smo tudi model, ki je kompatibilen s knjižnico in naučen na bazi podatkov. Ker globoko učenje za učenje potrebuje vzorec, ta pa mora biti pri prepoznavi slik še posebej velik (preko 100 000 slik), smo se odločili uporabiti bazo podatkov (*database*), ki ima vzorec slik že posnet in kategoriziran. Prav tako smo se odločili, da bomo uporabili model, ki se je na omenjenih slikah že naučil prepoznave novega materiala. Uporabili smo model *SSD_mobilenet_v1_COCO*, ki je bil naučen na bazi podatkov, ki smo jih uporabili.

Model deluje tako, da sliko najprej razdeli na $N \times N$ delov (celic) in za vsako celico predlaga 4 ali 6 možnih okvirčkov. V vsaki izmed celic, ki bi lahko vsebovala objekt, označi 4 (ali 6) okvirčkov, ki so v vseh primerih enaki. Nato določi, kateri okvirčki so najbolj primerni z uporabo vrednosti »SAMOZAVEST«. Postopek potem ponovi večkrat, in sicer v mreži z različnim številom in posledično različno velikostjo celic (conv layers). Velikosti so:

- 38*38
- 19*19
- 19*19 (6 različnih okvirčkov)
- 10*10
- 5*5
- 3*3
- 1

Velikosti si sledijo v tem vrstnem redu, saj model zaznava najprej manjše objekte in potem večje objekte, ki bi lahko bili na istem mestu. To je pomembno, saj lahko tako natančneje določi velikost objekta in dimenzije



Slika 5: Shema delovanja SSD algoritma (Sorokina, 2017)

okvirčka, ki objektu pripadajo. Na primer: če na nekem mestu zazna objekt, ko uporablja 38*38, in na istem mestu objekt, ko uporablja 19*19, in je na slednjem nivoju vrednost samozvesti večja, ve, da mora izbrati objekt z večjo samozvestjo, saj to pomeni, da se bo objektu okvirček lepše prilegal, saj je v danem primeru zelo verjetno, da okvirček v mreži 38*38 ne pokriva celotnega objekta. To pomeni, da bo ob konstantni ločljivosti izbral vedno večji objekt.

Algoritem vedno preveri, kateri kategoriji, ki mu je bila podana, ustreza objekt v okvirčku in s kakšno verjetnostjo lahko to trdi. Na koncu izpiše te rezultate, za katere je stopnja prepričanja za dano kategorijo relativno velika (odvisna od slike).

(Khandelwal, Towards data science, 2019) (Wei Liu, 2015)

3.3 Baza podatkov

Uporabili smo model, treniran na bazi podatkov **COCO** (Common Objects in COntext- običajni predmeti v kontekstu). Izbrali smo jo zaradi treh razlogov:

- Ta baza podatkov najbolj ustreza človeški prepoznavi objektov, saj se človek pri prepoznavi zanaša na okolje in prav tako večino prepoznav naredi v vsakodnevnem razgibanem okolju. To okolje predstavlja ozadje slik.
- Do vseh (80 kategorij) in tudi učnega materiala je mogoče prosto dostopati preko uradne spletne strani.
- Ta baza podatkov se najbolj ujema z modelom, ki smo ga uporabili, saj se po prejšnjih raziskavah model najbolj obnaša prav s to bazo podatkov. (Wei Liu, 2015)

Baza podatkov uporablja svoj API (application programming interface – vmesnik za aplikacije), ki omogoča, da z bazo podatkov, ki vsebuje več kot 330 000 primerov slik, model in algoritem najbolj učinkovito upravljata.

Model, ki smo ga uporabili, je bil torej že naučen, katere slike in objekti spadajo pod katero kategorijo, in je tako bil pripravljen sprejeti nove primere.

(COCO,
2019)
(Khandelwal,
Towards data
science, 2019)



Slika 6: Prikaz kategorij baze podatkov COCO (COCO, 2019)

4 METODOLOGIJA DELA

4.1 Izbor slik

Ker je bilo potrebno opraviti primerjavo med računalnikom in človekom, je bilo potrebno zbrati dva vzorca podatkov, ki jih je bilo treba primerjati na istem vzorcu slikovnega materiala. Pri izbiri slik smo natančno preučili vsebine slik in kot vzorec določili 15 slik, ki smo jih umestili v tri glavne kategorije, ki smo jih želeli preučiti. Prvo izmed kategorij (glej slike 1, 2, 3, 6, 9, 10, 14) lahko poimenujemo »osnovne slike«. V to kategorije spadajo slike, katerih elementi med seboj nimajo opaznih interakcij in slike, ki niso optične iluzije v pomenu, da se lahko v eni sliki razbereta dva pomena. V kategoriji so predvsem objekti, ki so postavljeni v vsakodnevni kontekst, z izjemo slik 1 in 2, ki sta objekta (mačka in pes), postavljena pred ozadje brez podrobnosti. Njun namen je postavljanje kontrolne osnove za prepoznavo, saj ne vsebujeta ozadja s kontekstom, ki je lahko za ljudi in računalnik moteče. Druga kategorija (glej slike 4, 13) je sestavljena iz slik, katerih elementi skupaj sestavljajo celoto ali opisujejo dejanje. Imenujemo jih »slike v kontekstu«, saj mora opazovalec ugotoviti, da se objekti na sliki v okviru nekega konteksta povezujejo. Tretjo kategorijo (glej slike 5, 7, 8, 11, 12, 15) sestavljajo slike, v katerih je moč videti dva objekta hkrati in so »optične iluzije«. Primeri v tej kategoriji so raznoliki, saj so tako barvni kot črno-beli, in različnih natančnosti. S to raznolikostjo smo želeli primerjati širino prepoznavanja objektov. Za to smo izbrali optične iluzije, ker gre za t. i. »mejne primere«, kjer se tudi človeški um ne more natančno odločiti, kaj vidi, in so tako odlični za primerjanje obnašanja računalnika v primerjavi s človekom, saj so na samem robu človeškega prepoznavanja. Vse pripadnike vseh treh kategorij smo med sabo pomešali.

4.2 Človeška metoda

Vzorci človeških osebkov smo pridobili s pomočjo uporabe strukturiranega intervjuja, kjer je sodelujoči ob PowerPoint predstavitvi z vzorčnimi slikami prepoznaval objekte. Intervju pomeni, da je pridobivanje podatkov potekalo ob prisotnosti intervjuvanega in voditelja intervjuja (enega izmed avtorjev). Intervju je bil močno strukturiran, kar pomeni, da so bila vprašanja vnaprej določena in da podvprašanja niso bila dovoljena. Močna strukturiranost intervjuja je pomenila tudi, da se je moral vprašani strogo držati navodil, ki so mu bila pred izvedbo podana. Intervju je potekal pod nadzorovanimi pogoji, na podobnih zaslonih in v okolju

z enakim nivojem motečih dejavnikov iz okolja. Vsem sodelujočim so podali enaka navodila. Z namenom čim večje nepristranskosti smo zastavili vprašanje: »Kaj vidite na sliki?«. Vprašanje se nam je zdelo najbolj nevtrarno, saj smo se želeli izogniti postavljanju vprašanj kot so: »Opišite sliko«, ali: »Kaj se dogaja na sliki?«, kar bi spremenilo odgovore tako, da bi bolj ustrezali določeni kategoriji, torej bili bolj pristranski. Prav tako nismo želeli slik ločiti v kategorije in zastaviti za vsako kategorijo posebnega vprašanja, saj bi tako intervjuvanci lahko predvideli kaj bo na sliki in podali pristranske odgovore. Da bi se izognili morebitnim nepravilnostim v primerjavi, je lahko vsak anketiranec na določeno sliko odgovoril po principu asociacij. Za vsako sliko je povedal le-to, kar se mu je ob prvem pogledu zdelo, da vidi na sliki. To pomeni, da dolgo razmišljanje ni bilo dovoljeno, kar pride v poštev še posebej pri slikah, ki imajo dva pomena, kjer bi lahko ljudje po razmisleku podali odgovor, ki se ne ujema z njihovim prvotnim opažanjem, računalnik pa bi v takem primeru odgovoril z odgovorom, ki se mu zdi najbolj očiten. Vsaki testni sliki je sledil čisto črn zaslon, katerega namen je pripraviti sodelujočega na naslednjo sliko na način, ki zmanjša vpliv prejšnje slike na prepoznavo. Vse odgovore smo posneli z diktafonom in jih nato po izvajanju ankete natančno prepisali v razpredelnico. Rezultate smo nato prevedli v angleški jezik in se s tem izognili sopomenkam (npr. kuža-pes). S tem smo podatke še bolje pripravili na primerjavo, saj so podatki računalnika prav tako v angleščini. Za obdelavo smo podatke prevedli nazaj v slovenščino. Intervju smo izvedli na **69** udeležencih.

4.3 Računalniška metoda

Za drugi del primerjave smo identično skupino slik analizirali v algoritmu za prepoznavo objektov na računalniku. Uporabili smo tehnologijo globokega učenja. Algoritmu smo slike podali po serijah po tri, da smo lahko z razpoložljivo računalniško močjo najbolj časovno učinkovito predelali podatke. Za vsako sliko je algoritem povedal, kaj se mu zdi, da je na sliki, in podal odstotek prepričanosti za podano kategorijo. Postopek smo nato ponovili **69-** krat, da je bilo število ponovitev enako številu intervjuvancev. Kasneje smo ugotovili, da je bilo to nekoristno, saj je računalnik vedno podal enake rezultate. Rezultate smo zapisali in shranili za nadaljnjo obdelavo in primerjavo s človeškim vzorcem.

5 REZULTATI

Vse pridobljene rezultate, računalniške kot tudi človeške, smo uredili in statistično obdelali. Analizo smo naredili za vsako sliko posebej, a bomo v nadaljevanju predstavili skupaj slike, pri katerih je bil rezultat preprost in zelo podoben (večinoma slike prve kategorije). Slike bodo predstavljene v enakem vrstnem redu, kot jih je analiziral računalnik, in so bile v intervjuju. Slike so v tem vrstnem redu tudi v prilogi. Grafikoni predstavljajo vsak tri slike, kot je na njih zapisano. Za ta način smo se odločili, da bi se ognili prekomerni količini grafikonov.

Preglednica 1: slika 1, 2, 3

slika	ljudje (št.)	računalnik (%)
1	mačka (69)	"cat" mačka (95)
2	pes (69)	"dog" pes (94)
3	skodelica (68), kava (1)	"cup" skodelica (95)

Računalnik je po 69 ponovitvah vedno dobil identičen rezultat. Rezultat računalnika je v tabeli označen z odstotki (%), prepričanja za nek objekt, ki je napisan ob okvirčku.

5.1 Slike 1, 2, 3

Prve tri slike so vse članice prve kategorije, kar pomeni, da je bila vsebina le en objekt, ki pa ni bil optična iluzija. Podatki za posamično sliko so razvidni v preglednici in grafikonu 1 (Preglednica 1: slika 1, 2, 3, Grafikon 1: slike 1, 2, 3).

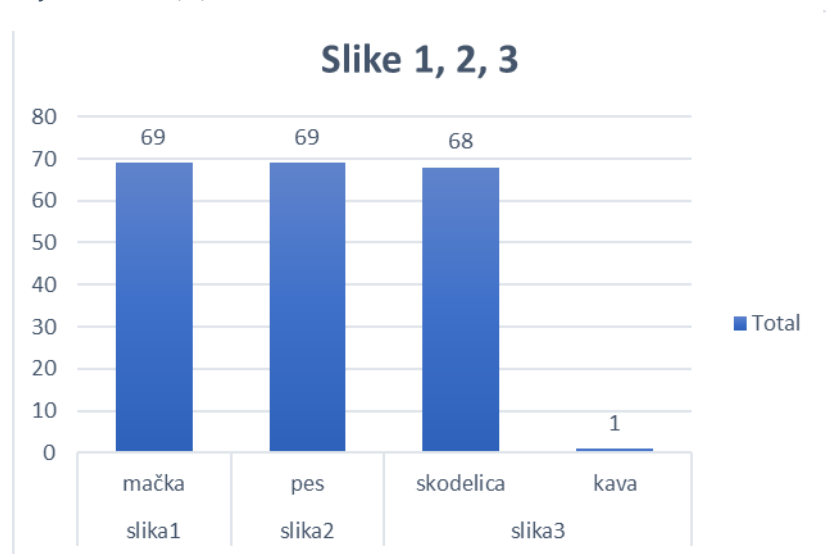
V sliki 1 je 100 % intervjuvancev prepoznalo mačko, rezultatu pa je sledil tudi računalnik, ki je vsebino označil za kategorijo »cat« (mačka) in 95 % (ostalih 5 % je razpršenih med ostalih 79 kategorij).

Pri sliki 2 je zgodba dokaj *Grafikon 1: slika 1, 2, 3*

podobna. Ljudje so se obnašali enako, saj skoraj vsi prepoznali psa. Računalnik je enako kot ljudje dodelil objektu oznako »dog« (pes), a je označil 94 % prepričanost.

Na sliki tri je 68 izmed 69 intervjuvancev prepoznalo skodelico, le eden je rekel, da

vidi kavo. Računalnik je s 95 % gotovostjo prepoznal »cup« (skodelico).



5.2 Slika 4

Slika 4 spada v drugo kategorijo (slike v kontekstu). Na sliki je prikazana cesta v mestu s parkiranimi avtomobili. Preizkušali smo, ali bodo ljudje ali računalnik znali prepoznati celoto in nasloviti sliko z »ulica«.

Kot je razvidno v preglednici 2 (Preglednica 2: slike 4, 5, 6), smo dobili precej razgibane rezultate. Po pričakovanjih je največ (39) posameznikov v sliki prepoznalo ulico, 16 vprašanih je videlo »mesto«, 6-krat pa so intervjuvanci prepoznali tako »cesto« kot »avto«. Ostali odgovori so bili razdeljeni med: »New York« in »ulica z avti«. Računalnik je na sliki prepoznal le dele, dvakrat je zaznal »car« (avto), in sicer z 79- in 76-odstotno prepričanostjo. V daljavi je z nižjo, 54-odstotno prepričanostjo, zaznal »person« (oseba). Po ogledu rezultatov je računalnik pravilno zaznaval in klasificiral.

Preglednica 2: slike 4, 5, 6

slika	ljudje (št.)	računalnik (%)
4	ulica (39), mesto (16), avto (6), cesta (6), new york (1), ulica z avti (1)	"car" avto (79, 76), "person" oseba (54)
5	telo (62), 2 kroga (1), ženska (3), 2 osebi, ki plešeta (1), N.G (1), prijatelja, ki si mahata ¹ ,	prazen okvir (/)
6	stol (69)	"chair" stol (89)

5.3 Slika 5

Kot prva predstavnica tretje kategorija optičnih iluzij je slika 5 prinesla zanimive rezultate. Sliko je možno razumeti na dva načina. Vidimo lahko žensko telo ali dva človeka, ki plešeta.

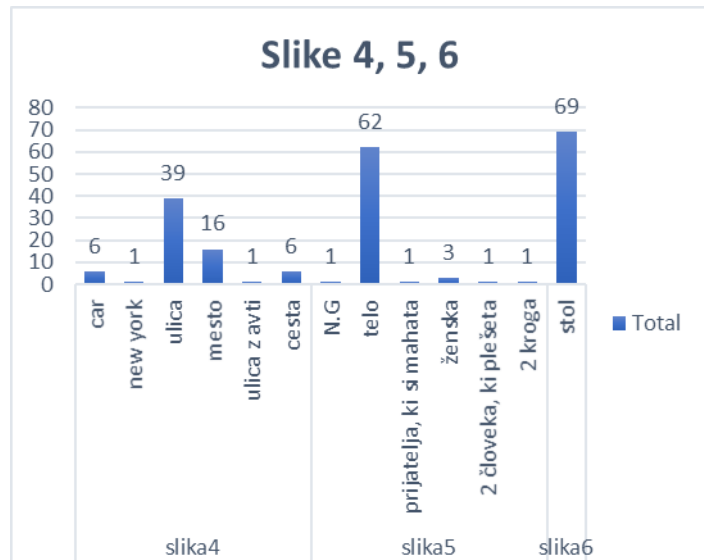
Odgovori pri človeku so bili nagnjeni

Grafikon 2: slike 4, 5, 6

proti prvemu pomenu (žensko telo), saj je 62 udeležencev videlo »telo«, trije pa »žensko«. En odgovor je bil »človeka, ki plešeta«, en vprašani kandidat pa je videl »prijatelja, ki si mahata«. Ta odgovor smo prav tako šteli kot drugi odgovor, saj je posameznik videl človeški figuri. Eden izmed intervjuvancev je videl dva kroga, eden pa odgovora ni podal, kar prikazuje grafikon 2

(Grafikon 2: slike 4, 5, 6).

Računalnik je v sliki uspel identificirati dva verjetna objekta, a jima ni uspel določiti kategorije.



5.4 Slika 6

Slika 6 je prav tako predstavnik prve kategorije. Med odgovori ljudi je vseh 69 intervjuvancev na sliki prepoznalo stol.

Računalnik je enako kot ljudje objekt označil s »chair« (stol) ter zaznavi pripisal 89 % natančnost.

5.5 Slika 7

Slika 7 je optična iluzija, v kateri lahko posameznik vidi podeželsko sceno in tudi obraz starejšega moškega.

Največ intervjuvancev (27) je v sliki prepoznalo »obraz«. Drugih 12 je trdilo, da na sliki vidijo »moškega«, eden je videl »glavo«. Oboje štejemo pod drugi pomen, ki ga je videlo skupno 40 sodelujočih. Drug pomen (podeželje) je prepoznalo ostalih 29 udeležencev. Največ jih je videlo ovce (10).

Tabela 4 prikazuje človeške odgovore za dotično sliko.

Računalnik v sliki ni uspel podati odgovora v nobeni od ponovitev.

5.6 Slika 8

Slika osem je prav tako optična iluzija, v kateri je moč videti ustnice ali drevje.

V grafu 3 (Grafikon 3: sliki 7, 8) vidimo, da se je za drevo opredelilo največ ljudi, skupno kar 43. Preostalih 26 je videlo ustnice.

Računalnik tudi pri tej sliki ni uspel postaviti predvidevanja.

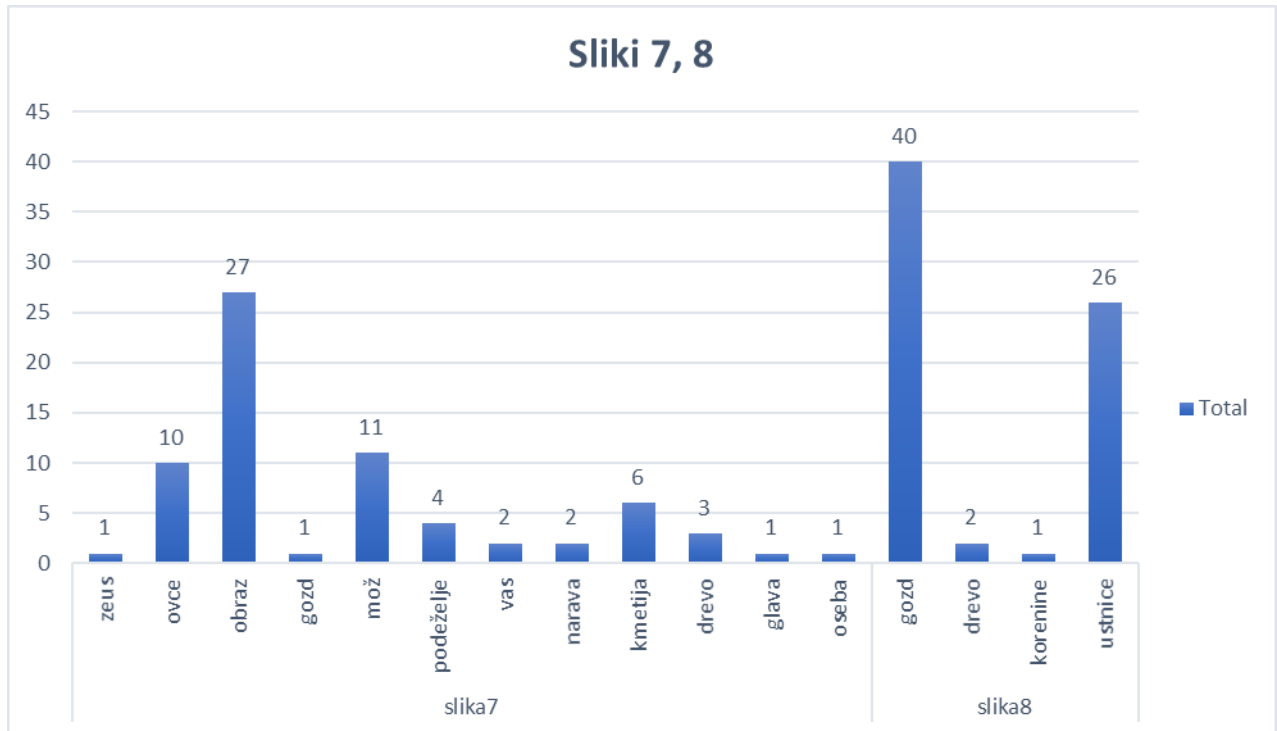
Preglednica 3: sliki 7, 8

Row Labels	Count of ID
slika7	69
zeus	1
ovce	10
obraz	27
gozd	1
mož	11
podeželje	4
vas	2
narava	2
kmetija	6
drevo	3
glava	1
oseba	1
slika8	69
gozd	40
drevo	2
korenine	1
ustnice	26

Preglednica 4: sliki 7, 8

slika	ljudje (št.)	računalnik (%)
7	obraz (27), ovce (10), kmetija (6), podeželje (4), vas (2), gozd (1), glava (1), mož (11), narava (2), oseba (1), drevo (3), zeus (1)	// (/)
8	ustnice (26), gozd (40), drevo (2), korenine (1)	// (/)

Grafikon 3: sliki 7, 8



Preglednica 5: slike 9, 10, 11, 12

slika	ljudje (št.)	računalnik (%)
9	letalo (69)	"airplane" letalo (97)
10	avto (67), mclaren (1), porche (1)	"car" avto (91)
11	drevo (37), opica in lev (20), živali (5), opica (3), lev (2), ribe (1), opica, tiger, drevo (1)	// (/)
12	don kihot (16), mlin na veter (10), konj (8), obraz (7), oseba (8), star mož (4), jezdec (5), človek (2), einstein (2), sancho pansa (2), vitez na konju (1), vitezi (1), človek na konju (1), jezdec (2)	"person" oseba (58)

5.7 Sliki 9 in 10

Obe sliki sta uvrščeni pod prvo kategorijo.

Na sliki 9 so vsi udeleženci prepoznali letalo. Letalo je prepoznal tudi računalnik, in sicer s 97 % gotovostjo.

Na sliki 10 je večina ljudi prepoznala »avto«, vsi so se pa opredelili za pojem avtomobil. Računalnik je prepoznal avto z 91 % gotovostjo.

5.8 Slika 11

Slika 11 je optična iluzija in ima dva možna pomena. Drevo in živali (lev, opica, ribe).

Za prvi pomen se je opredelilo 37 udeležencev. Za drugi pa skupno 32. Od tega je eden od intervjuvancev videl vse tri živali, eden pa le leva in opico.

Računalnik ni uspel najti objektov.

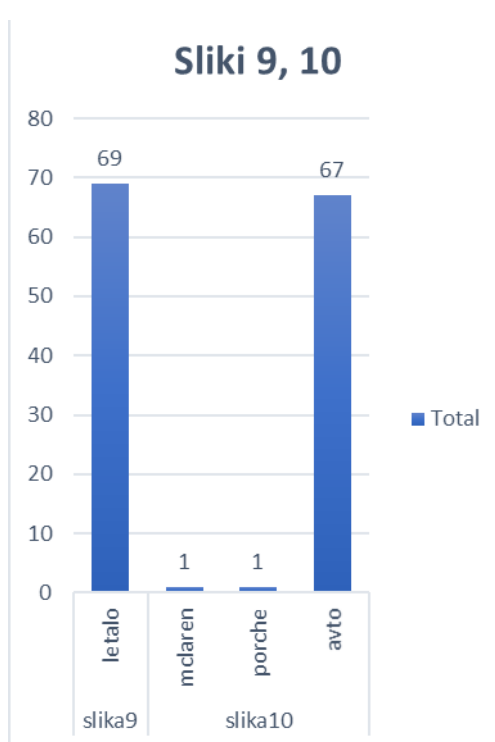
5.9 Slika 12

Slika 12 je optična iluzija, katere podatki so se izkazali za najbolj zanimive. Na sliki je bilo moč prepoznati obraz starejšega moškega (Einstein) ali sceno iz romana Don Kihot. Rezultati intervjuja so predstavljeni v grafikonu 5 (Grafikon 5: sliki 11, 12).

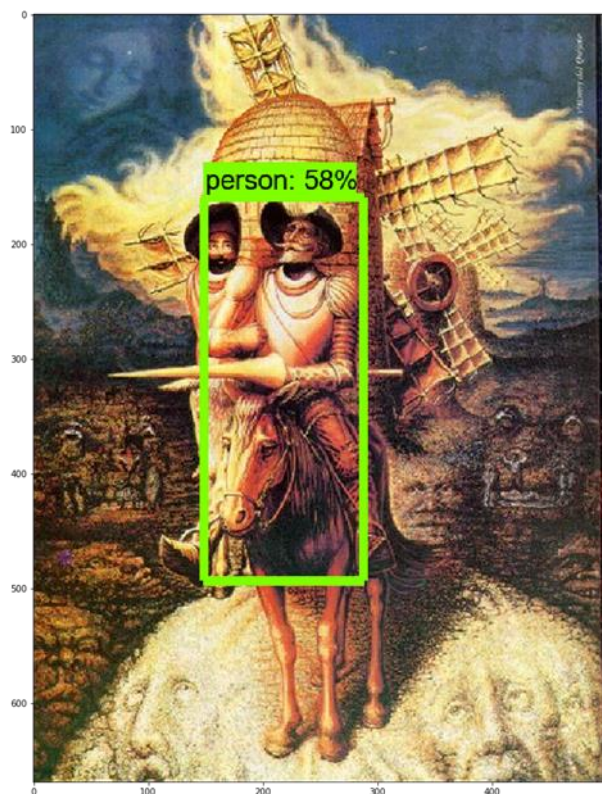
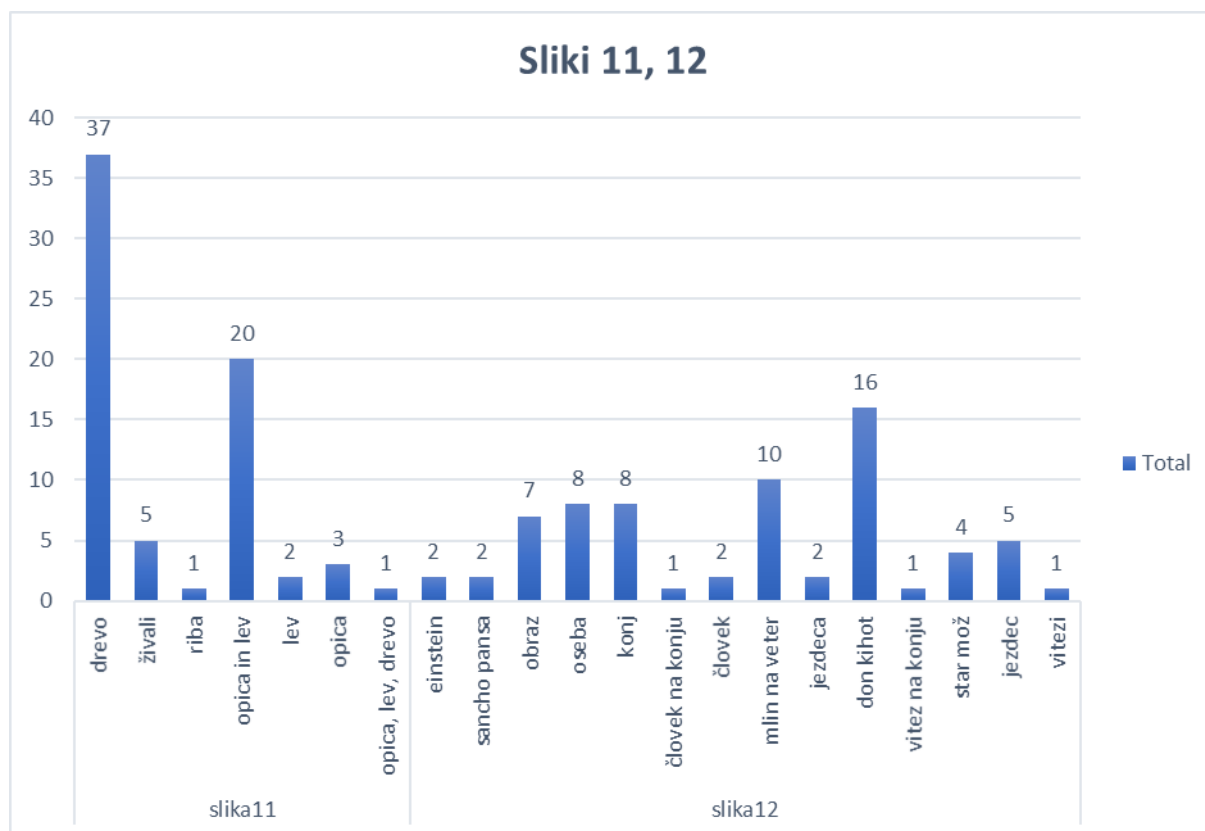
Obraz je skupno prepoznalo 13 udeležencev, sceno iz Don Kihota pa skupno 46 udeležencev. Ostalih 10 udeležencev je podalo odgovor, ki ga ne moremo umestiti pod dotičen pomen.

Računalnik je na sliki opazil osebo in okvirček narisal tako, da je prepoznal del scene z Don Kihotom. Zaznal je z 58 % natančnostjo.

Grafikon 4: sliki 9, 10



Grafikon 5: sliki 11, 12



Slika 7: Računalnikov rezultat slike 12 (Ocampo, 2020)

5.10 Slika 13

Slika 13 je pripadnica kategorije slik v kontekstu. Kontekst je v tem primeru »hranjenje«.

25 udeležencev intervjuja je sliko opisalo s kontekstom in prepoznalo neko vrsto hranjenja. Dva sta prepoznala »hrano«. Kar 37 je v sliki videlo posamezne komponente (človeka, psa). Trije so podali odgovor »skrbnik«, dva pa sta prepoznala globlje pojme (»ljubezen«, »prijateljstvo«).

Računalnik je razbral posamezne komponente slike. Identificiral je psa (96 %), osebo (93 %) in tudi nekaj predmetov v ozadju: kavč (60 %) in rastlina (50 %).

Preglednica 6: slike 13, 14, 15

slika	ljudje (št.)	računalnik (%)
13	pes (19), hranjenje (20), človek in pes (14), človek, ki hrani psa (5), človek (4), skrbnik (3), hrana (2), prijateljstvo (1), ljubezen (1)	"dog" pes (96), "person" oseba (93), "couch" kavč (60), "potted plant" rastlina (50)
14	Hamburger (69)	"sandwich" sendvič (95)
15	ptica (33), raca (18), zajec (16), raca in zajec(2)	// (/)

5.11 Slika 14

Slika 14 je preprosta slika prve kategorije, ki prikazuje hamburger.

Vsi intervjuvanci so na sliki prepoznali hamburger.

Računalnik je na sliki prepoznal sendvič, in sicer s kar 95 % natančnostjo.

5.12 Slika 15

Slika 15 je klasičen primer iluzije raca ali zajec.

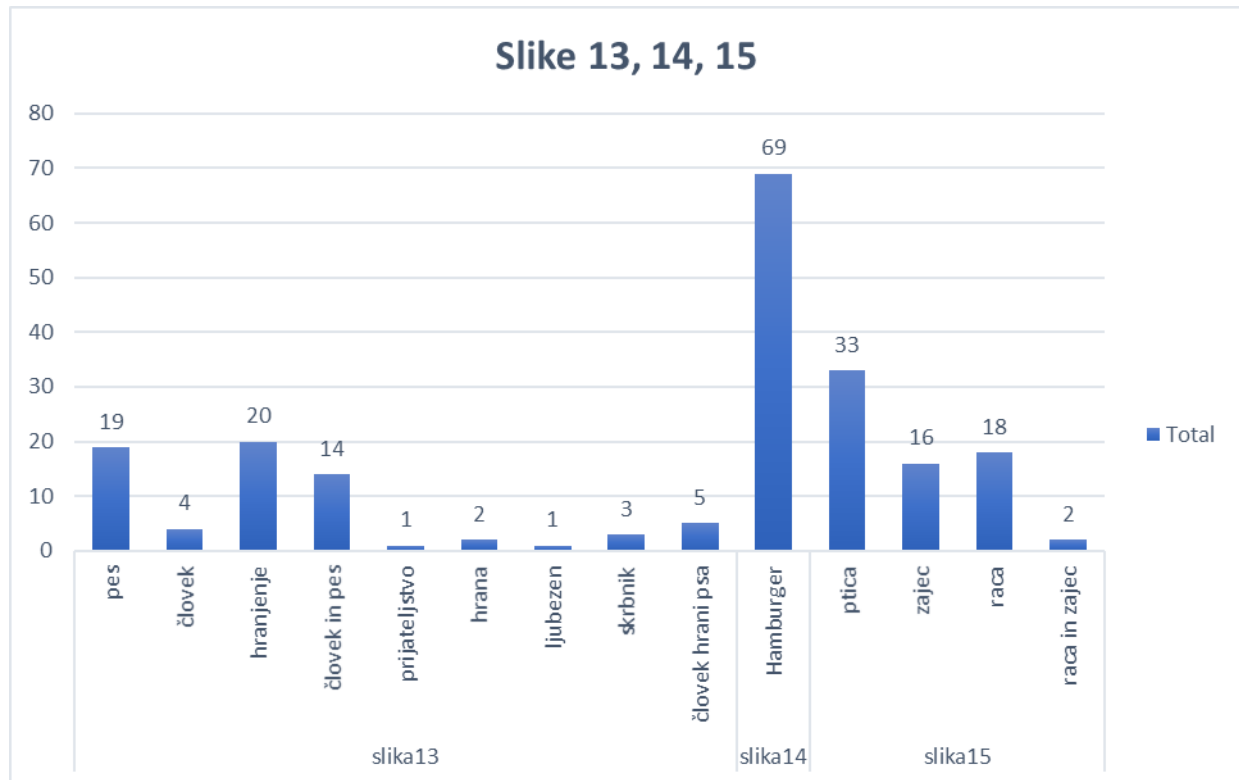
51 intervjuvancev je prepoznalo raco (ptiča), 16 zajca, dva pa sta na sliki takoj videla oboje.

Računalnik na sliki ni uspel najti objekta.

Preglednica 7: slike 13, 14, 15

Row Labels	Count of ID
slika13	69
pes	19
človek	4
hranjenje	20
človek in pes	14
prijateljstvo	1
hrana	2
ljubezen	1
skrbnik	3
človek hrani psa	5
slika14	69
Hamburger	69
slika15	69
ptica	33
zajec	16
raca	18
raca in zajec	2

Grafikon 6: slika 13, 14, 15



6 ZAKLJUČKI

6.1 Kategorija 1

Kategorijo 1 so sestavljale slike, na katerih naj bi se po naših predvidevanjih tako računalnik kot ljudje obnašali najboljše. Slike predstavljajo preproste objekte, ki imajo en pomen, ki najbolj izstopa. Primer take slike je »pes« na sliki 2. Na sliki je jasno razvidno, kaj je glavni subjekt in kaj je ozadje. Glavni subjekt je izpostavljen, medtem ko je ozadje docela nepomembno in v večini primerov celo nima vsebine. Izpostava je narejena z zamegljevanjem ozadja ali preprosto z uvedbo enobarvnega ozadja, kot je razvidno na primeru slik 1 in 3. Dejstvo, da je subjekt slike v tej kategoriji močno izpostavljen, je pomembno za analizo in razpravo o rezultatih. Slike so oštevilčene z 1, 2, 3, 6, 9, 10, 14.

Ljudje so se na večini slik strinjali glede rezultatov. Velika večina jih je izbrala enak rezultat. Največja raznolikost je bila še vedno 97 % enaka. Ljudje so se torej v večini primerov strinjali z vsebino odgovora, ki pa je bila tudi zmeraj pravilna. To je bilo pričakovano, saj ljudje prepoznavamo preproste objekte že od otroštva. Naša prepoznavna deluje zelo učinkovito in natančno, ko je objekt jasno določen in izpostavljen na sliki ter ozadje ni moteče.

Računalnik se človeški prepoznavi v tej kategoriji izjemno približa. To je najverjetneje mogoče zato, ker je računalnik treniran, da zazna objekte, ki s slike izstopajo oziroma, da zazna objekte, katerih pomena in kategorije ozadje in drugi moteči dejavniki ne zabrišejo preveč. Prav tako je baza podatkov naučena na primerih, ki so ločeni v kategorije, ki ponazarjajo primere objektov iz vsakodnevnega življenja. Tako računalnik brez težav najprej odkrije, kje je objekt (s pomočjo preprostega ozadja), nato pa lahko v okvirčku, ki ga je določil, brez težav pripiše kategorijo, za katero je bil predhodno že naučen s pomočjo tisoč podobnih primerov, ki v modelu povežejo nevrone, da ob istih značilnostih dajo enak rezultat.

Pri teh slikah je bila primerjava uspešna, zato lahko trdimo, da bi se računalnik v okolju, ki bi od njega zahteval le preprosto prepoznavo, vedel kot človek, oziroma da zunanji opazovalec ne bi razločil, ali je sliko prepoznaval človek ali pa računalnik.

Pri slikah 1, 2, 6 in 9 je bil objekt posebej izrazito izpostavljen, saj je ozadje preprosto bele barve ali zelo zamegljeno. Objekt ima prav tako veliko podrobnosti, kar dodatno pomaga pri prepoznavi. Prepoznavne slik so imele 100 % enotnost med ljudmi, računalnik pa je imel 95 %,

94 %, 89 % in 97 % prepričanost. Vedeti je potrebno, da računalnik vedno preveri vse kategorije (v našem primeru 80) in jih 100 % razporedi po prepričanju. Tako računalnik nikoli ni 100 % prepričan oziroma se to zgodi zelo redko. To je ena pomembnejših razlik med človekom in računalnikom, saj se človek običajno odloči za eno stvar, ostale možnosti pa zavrne.

Slika 3 od slike 1 loči ena stvar. Eden od 69 udeležencev intervjuja je na sliki videl »kavo« čprav je bila prikazana prazna skodelica. To kaže, da lahko ljudje na sliki prepoznajo nekaj, kaj jih spominja na drug pojem – asociacija.

Zanimivost pri sliki 10, ki prikazuje avto, izhaja iz človeškega zaznavanja. Dva izmed sodelujočih sta na sliki avta namreč prepoznala znamke avtov. Prepoznala sta Porsche in McLaren, čprav je bil avto Porsche. To kaže, da kultura in civilizacija vplivata na našo prepoznavno sposobnost in na naše asociacije. Odgovor McLaren je morda bil podan pod vplivom osebne preference.

Računalniška prepoznavna je zgrešila le pri sliki 14, ki prikazuje hamburger. Za imena »sendvič« je krivo pomanjkanje kategorij oziroma močna podobnost med objekti, zaradi katere bi bilo narediti dve ločeni kategoriji nesmiselno. Računalnik je tako prepoznal nekaj, kar zgleda zelo podobno in je v bistvu enaka stvar, a smo ljudje v naši kulturi raje ločili pojma »hamburger« in »sendvič«. Ljudje so seveda poznali obe kategoriji. Vseeno pa lahko rečemo, da lahko računalnik, kljub temu da je kategorija malenkost drugačna, vrne rezultate, ki so pomensko enaki človeškim.

6.2 Kategorija 2

Pri kategoriji 2 je šlo za kategorije slik s kontekstom ali z dogajanjem. Pri tej skupini smo uporabili slike, ki so vsebovale nekakšno dogajanje med objekti. Imajo pa tudi več različnih objektov, kot je na primer slika 13, »hranjenje«. Na tej sliki vidimo dejavnost med človekom in psom, ki je hranjenje. V tej kategoriji ne gre samo za en prikazan objekt v glavni vlogi, ampak za več teh hkrati, ti pa skupaj tvorijo neko novo asociacijo za človeka, kar je pomembno pri analizi rezultatov. Med te spadata slika 4 in 13.

Ljudje so pri tej kategoriji zaznali neko celoto oziroma aktivnost med objekti. Proti pričakovanjem je širše dogajanje zaznalo le 50 % intervjuvanih. Prepoznali so ali ulico ali hranjenje, ostali anketiranci pa so prepoznali posamezne elemente na sliki. Naša pričakovanja so bila napol uresničena. Nismo si mislili, da bo velika večina sicer prepoznala posamezne

objekte, vendar jih ne bo povezala s celoto. Možnih je več vzrokov, eden od njih je tudi ta, da ljudje različno interpretirajo podana navodila. Zanimiva odgovora pa smo dobili tudi pri sliki 13. Eden izmed 69 udeležencev ankete je prepoznal ljubezen, eden pa prijateljstvo. Takšne odgovore lahko pričakujemo samo pri ljudeh, in ne računalnikih, saj se s tem pokaže človeška čutnost in prepoznavanje čustev. Medtem ko je računalnik naučen, da zazna objekte, kot so človek in pes posebej in jih da vsako v svojo kategorijo, lahko človek najprej vidi smeh na drugi osebi in ga takoj poveže s prijetnimi čustvi, kot so ljubezen, veselje, prijateljstvo. Pri sliki 4 je prišlo do nekaj različnih odgovorov. Ljudje so, sodeč po anketi, najprej prepoznali ulico, mesto, avto, New York, cesto in ulico z avti. Do toliko različnih odgovorov je najverjetneje prišlo zato, ker je na sliki veliko različnih objektov, tako v ospredju kot v ozadju. Na njihove odgovore je najbolj vplivalo, kateri del ekrana so gledali, ko se je prikazala slika, in kakšno je njihovo preteklo življenje oziroma s čim so se največkrat srečali, kar povzroča del seznanjenosti pri človeški prepoznavi, torej kvantitativni podatki. Tako je na primer oseba, ki je živela oziroma živi v podobnem okolju in doma temu pravijo mesto, odgovorila z »mesto«, medtem ko je oseba, ki je odraščala v okolju, kjer so to imenovali ulica, odgovorila z »ulica«. Opazili pa smo tudi odgovor »New York«, kar si lahko razlagamo na dva načina. Ali je ta oseba skozi življenje bila obkoljena z mestom New York in ji vsaka večja ulica priključa asociacijo na Ameriko, ali pa ni želela odgovoriti resno, kar se pri ljudeh lahko zgodi. Pri odgovorih, ki potrebujejo več asociacij, sta delovala oba procesa prepoznavne, tako spomin kot seznanjenost, medtem ko so ljudje, ki so odgovorili z »avto« ali s »cesta«, uporabljali samo proces seznanjenosti.

Dejstvo, da so ljudje pred "ulica" izbrali "avto", kaže na to, da se človeški um v situacijah, ko je potrebno uporabiti seznanjeno prepoznavo, fokusira na ožji objekt namesto na širšo sliko. To je lahko posledica več stvari. Lahko je krivo dejstvo, da želimo poenostaviti sliko in gledati le ozek del ali pa je razlog biološki. Možno je, da zaradi ozkega vidnega polja (rumene pege) ob danem trenutku vidimo le majhen del slike.

Računalnik je s pomočjo algoritma pri tej kategoriji deloval oziroma podal odgovore tako, kot smo pričakovali. Pri slikah, ki imajo več objektov in predstavljajo celoto oziroma dejanje, računalnik prepozna samo posamezne objekte in jih uvrsti med kategorije, v našem primeru 80, medtem ko človek prav tako prepozna posamezne objekte, a jih zna povezati v celoto. To lahko opazimo pri sliki 4, kjer računalnik prepozna dva avta z 79- in 76-odstotno prepričanostjo in človeka v daljavi s 54-odstotno prepričanostjo. Vidimo lahko, da prepozna več različnih elementov hkrati, a jih ne zna povezati v skupek, ki bi tvoril celotno sliko. Pri sliki 13 računalnik

prepozna psa, osebo rastlino in kavč. Računalnik lahko naenkrat oziroma v zelo kratkem času pregleda celotno sliko in ugotovi, kaj vse vidi, pri človeku pa opazimo, da nobeden od udeležencev ni odgovoril z »rastlina«, saj se ta pojavi v ozadju in ni čisto izostrena. Kot taka ne pritegne pozornosti, zato človek najprej opazi tisto, kar je v osredju, medtem ko lahko računalnik naenkrat preuči celotno sliko. Računalnik tako opazi stvari v ozadju.

Primerjava je zanimiva, saj ljudje niso prepoznali konteksta v tolikšni meri, računalnik pa je, čeprav je prepoznal objekte, opazil stvari, ki jih človeški opazovalec ni. Vidimo tudi, da pri računalniku prepričanost v nek objekt pada, kar je posledica vedno manjše ločljivosti. Podobno bi se lahko zgodilo pri človeku, a se človek, ko najde nek pomen, zadovolji s tem in ne raziskuje ozadja.

6.3 Kategorija 3

V tretji kategoriji so bile slike, v katerih naj bi človeški opazovalci videli dva različna objekta v odvisnosti od perspektive. Najboljši primer je slika 15, na kateri lahko vidimo raco ali zajca. V kategoriji sta se ustvarili dve nekakšni podkategoriji, ena s slikami, ki so črno-bele, druga pa z barvnimi slikami. Rezultati v teh kategorijah so malce različni, še posebej pri računalniški prepoznavi. Slike so ilustracije, kar je eden od možnih razlogov za slabšo računalniško prepoznavo. Računalnik je namreč uspel kategorizirati objekt le pri eni izmed slik (12), pri še eni pa je uspel zaznati, da je na sliki objekt (5). Na slikah se dva pomena lahko razpoznata na dva glavna načina. Pomen je torej pogojen z orientacijo objekta (en objekt ima dva pomena v odvisnosti od orientacije), kar lahko vidimo pri sliki 15, ali pa je mogoče dva pomena razpoznati s tem, da gledamo ožjo ali širšo sliko, kar je možno opaziti pri sliki 12. Slike kategorije 3 so: 5, 7, 8, 11, 12, 15.

Človeški odgovori so bili v tej kategoriji bolj razgibani, kot je bilo sprva pričakovano. Pri mnogih slikah se odgovori niso delili le v dve kategoriji, ampak so bili odgovori razgibani tako v besednem zakladu kot tudi v opažanju več pomenov slike hkrati. Vse odgovore se je dalo brez težav opredeliti v eno ali drugo kategorijo, a so bili zelo nekonsistentni. To se pozna še posebej pri barvnih slikah, ki imajo, da dosežejo dvopomenskost, veliko detajlov, in je zato možno, da posameznik opazi enega izmed teh, kar vodi do širše palete odgovorov. To je razširjeno do te mere, da pri nekaterih slikah sploh ni kategorije, ki bi zajemala več kot 50 %. Do neke mere vidimo to raznolikost tudi pri črno-belih slikah, čeprav ni niti približno tako izrazita in obširna kot pri barvnih. Pri teh se zaradi manj detajlov udeleženci opredelijo v eno ali drugo lastnost,

ali pa vidijo nek pojem, ki ga nismo pričakovali, kar je najverjetneje posledica osebnih asociacij posameznika.

Računalniška prepoznavna v tej kategoriji ni izpolnila pričakovanj, saj je računalnik objekt našel le v dveh izmed šestih slik, kategoriziral pa je objekt le v eni. Primera bosta v nadaljevanju obdelana. Za večino slik računalnik najverjetneje ni našel objekta, ker je šlo za ilustracije, ki imajo zelo malo skupnega z objekti, ki naj bi jih predstavljale v resničnem življenju. Računalnik tako ne najde zadostne verjetnosti, da bi neko območje klasificiral kot objekt, kaj šele da bi ga kategoriziral. Pri kategorizaciji pa je edina slika s pozitivnim rezultatom barvna, kar kaže na to, da barvne slike bolje delujejo, kar je povsem logično, saj je bil model naučen na primerih iz resničnega življenja, ki so v barvah.

Primerjalno gledano je v tretji kategoriji vse obrnjeno v smer ljudi, saj računalnik niti približno ni uspel dohajati^[TM2]. Računalnik namreč, ko nima dovolj konteksta, prepričanja ali možnih mest za objekt, ne poda kategorije, ljudje pa se pri interpretaciji takšnih slik odločimo za nekaj. Vredno je omeniti, da računalnik nima toliko podatkov, kot jih imamo ljudje, saj se učimo daljše obdobje in imamo, čeprav potrebujemo več časa, več vzorčnih podatkov. Model prav tako ni naučen za prepoznavanje ilustracij in črno-belih slik, kar pa ne pomeni, da to ne bo nikoli možno med prosto dostopnimi modeli.

Slika 5 je prvi zanimiv primer tretje kategorije. Slika lahko prikazuje preprosto skico ženskega telesa ali pa preprosto skico dveh plesalcev. Na sliki je računalnik uspel identificirati možen objekt, a ga ni uspel klasificirati. Zanimivo je dejstvo, da okvirček pokriva celotno sliko, ki predstavlja telo, kar se presenetljivo ujema s človeško zaznavo. To pomeni, da računalnik na sliko gleda podobno kot ljudje (prednost daje večjemu objektu, če sta oba objekta prikazana z enako ravno podrobnosti). To je posledica načina delovanja algoritma. Ta namreč pri razdeljevanju na $N \times N$ mreže daje prednost večjim mrežam in večjim objektom. To pa ne pomeni, da manjši, izrazitejši objekti ne dobijo prepoznavne. Izjemen primer tega pojava je slika 12.

Na sliki 12, kljub temu da je upodobljen večji obraz starejšega gospoda, računalnik in ljudje prepoznajo osebo na konju, kar kaže na to, da oboji izberejo manjši objekt, če je le-ta bolj izrazit in prej pade v oči. To kaže na sposobnost računalnika, da podobno kot ljudje daje prednost izrazitejšim objektom preden pogleda velikost. Bolj podobna kot je izrazitost, večji vpliv ima velikost.

7 ZAKLJUČEK

V nalogi smo raziskali računalniško prepoznavo objektov in jo primerjali s človeško. Prišli smo do zanimivih ugotovitev. Najprej je vredno povedati, da se tehnologija, še posebej na področjih strojnega in globokega učenja, razvija z nepredstavlljivo hitrostjo in je zato zelo verjetno, da so mnoge napake, ki smo jih odkrili, že v procesu popravljanja.

Ugotovili smo, da je računalnik izjemno podoben človeku. Pri prepoznavanju preprostih slik je na človeku enakem nivoju. Zna namreč določiti pravilno kategorijo, in to z relativno veliko natančnostjo. Tako lahko rečemo, da je naša prva hipoteza **potrjena**. Z gotovostjo lahko rečemo, da je računalnik, čeprav njegova gotovost ni 100-odstotna (kar je posledica delovanja), zadosti dober, da je na človeškem nivoju.

Računalniška prepoznavna pa se od človeške loči še v drugi kategoriji. Računalnik ne zmore prepoznati konteksta (npr. ulica), določi lahko le njene dele. Manjše presenečenje pa je ravnanje ljudi. Tudi ljudje so v večini prepoznali le dele slike, le manjšina je povzela celoten kontekst. Tako lahko rečemo, da je druga hipoteza **delno potrjena**, saj se je računalnik obnašal pričakovano, človeški odgovori pa so bili razpršeni.

Glavna ugotovitev tretje kategorije je, da računalnik ni zmožen prepoznave tako širokega spektra primerov, kot smo sprva mislili. Ker v večini primerov ni uspel prepoznati ničesar, moramo kljub obetavni sliki 12, hipotezo **zavreči**, saj preprosto ne moremo trditi, da na tem področju računalnik drži korak s človeškim umom.

Ugotovili smo, da je računalniška tehnologija napredovala dovolj, da lahko v praktičnih delih nadomešča človeka. Vsa dela, kjer je treba uporabiti preprosto prepoznavo (samovozeči avtomobili, sortiranje slik, sortiranje objektov), so za računalnik neproblematična. Ko pa pride do tega, da mora računalnik delovati v okolju in povezati objekte, ki jih vidi, v smiselno celoto, tega na tem nivoju še ni sposoben.

V prihodnje se zdi pomembno raziskati, kaj je povzročilo pomanjkanje računalniških odgovorov pri optičnih iluzijah. Ali je pomanjkljivost v ilustracijski naravi slik ali v pomanjkanju barve. Ali pa je krivo dejstvo, da so slike zaradi dvopomenskosti čudne za prepoznavo, kot so čudne ljudem. Zanimivo bi bilo opraviti raziskavo, ki bi preverila, kako se računalnik obnaša, ko spreminjamo te tri spremenljivke, in določiti, kje je pomanjkanje.

Zanimivo bi bilo opraviti podobno raziskavo z uporabo več različnih modelov in algoritmov tudi na večjem številu slik. Pravzaprav bi povečanje razsežnosti pomagalo povsod. Z vpletanjem več raziskovalcev bi lahko pokrili večjo populacijo ljudi in raziskavo komentirali z več zornih kotov. Pomagalo bi tudi (še posebej pri drugi kategoriji), da bi eksperiment na ljudeh opravili v naravnem okolju.

Vsekakor je napredek tehnologije izjemno hiter in raziskavo bi se dalo v prihodnosti z dostopnostjo vse več odprtokodnih algoritmov in modelov zelo razširiti. Vseeno se porajajo okoli te relativno mlade tehnologije mnoga vprašanja in mnogo predlogov za izboljšavo. Eden od teh je dejstvo, da so spektri prepoznave relativno ozki (prepoznavna vsakodnevnih objektov, prepoznavna števka ...). Vsa področja prepoznavne, ki so v človeških možganih združena, bi lahko združili tudi v računalniku in tako omogočili celovito prepoznavo.

Vsekakor je bil namen naloge dosežen. Ugotovili smo, kakšne razsežnosti lahko ima računalniška prepoznavna, ki je prosto dostopna vsakomur, v primerjavi s človekom. Določili smo, kako dobro se obnese na uporabnih področjih in kako deluje na meji človeške prepoznavne in kako pri optičnih iluzijah.

8 DRUŽBENA ODGOVORNOST

Družbena odgovornost je ena od stvari, ki jih mora znanost upoštevati, če želimo svet oblikovati v smeri, ki je dobra za človeštvo.

Ta naloga se posveča temi, ki je trenutno zelo aktualna – umetna inteligenca (U. I.). Tehnologija vse hitreje napreduje in področje umetne inteligence ne zaostaja za razvojem. Kmalu bo potrebno razmišljati o vključevanju U. I. v družbo, v kateri živimo. Ta naloga odpira debato v smeri razmisleka o vse večjem približevanju U. I. človeškemu umu. Zato je zelo pomembno, da ljudi ozaveščamo o njenem pomenu in točki v razvoju, saj je mnoge ljudi strah, da bo umetna inteligenca kmalu prevladala nad ljudmi. Tako je potrebno že zdaj začeti razmišljati, kako jo družbeno odgovorno uvesti.

Vidik naloge je aktualen in zato lahko vpliva na bralca. Želimo povedati, da računalniki še niso na točki, da bi lahko zamenjali ljudi, a se temu vztrajno približujejo. Za oblikovanje prihodnosti, katere del je in bo tudi U. I., je pomembno, da se vsi zavedamo, kako dotična tehnologija napreduje in kako bo vplivala na družbena razmerja. Vsi bralci naloge in udeleženi imajo tako neposredno možnost sodelovati v razpravi o razmerjih med človekom, družbo in U. I.

Ker je možno, da bo U. I. prevzela mnoge funkcije, ki jih sedaj opravljajo ljudje, je že zdaj potrebno, da se uredi implementacija U. I. v družbo in da se začne debata in razprava o U. I., kar bo omogočilo trajnostni razvoj v družbi prihodnosti, katere člani morda ne bodo le ljudje.

9 PRILOGE

9.1 Priloga 1 – slike, sodelujoče v raziskavi

Slika 1: (Wikimedia Commons, the free media repository, 2016) Slika 2: (Cohut, 2018)



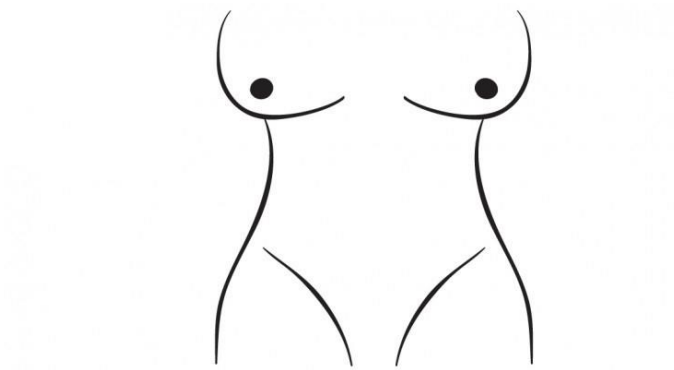
Slika: 3 (Wikimedia Commons, the free media repository, 2016) Slika 4: (Wikimedia



Commons, the free media repository, 2019)



Slika 5: (Abrozzi, 2020)



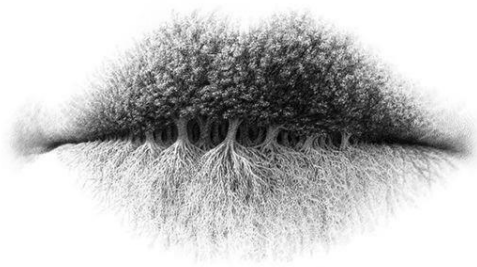
Slika 6: (Wikimedia Commons, the free media repository, 2018)



Slika 7: (Pinterest, 2020)



Slika 8: (Dagorov, 2020)



Slika 9: (Wikimedia Commons, the free media repository, 2017) Slika 10: (Wikimedia Commons, the free media repository, 2019)



Slika 11: (Anderton, 2020)



Slika 12: (Ocampo, 2020)



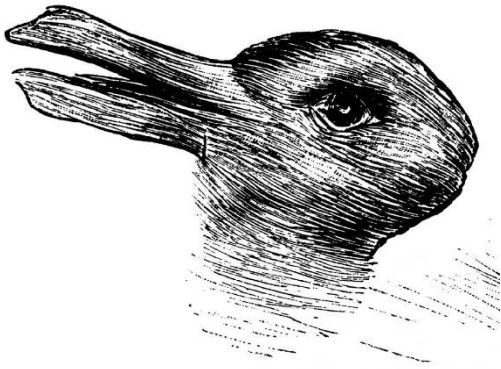
Slika 13: (Petside, 2019)



Slika 14: (Wikimedia Commons, the free media repository, 2018)

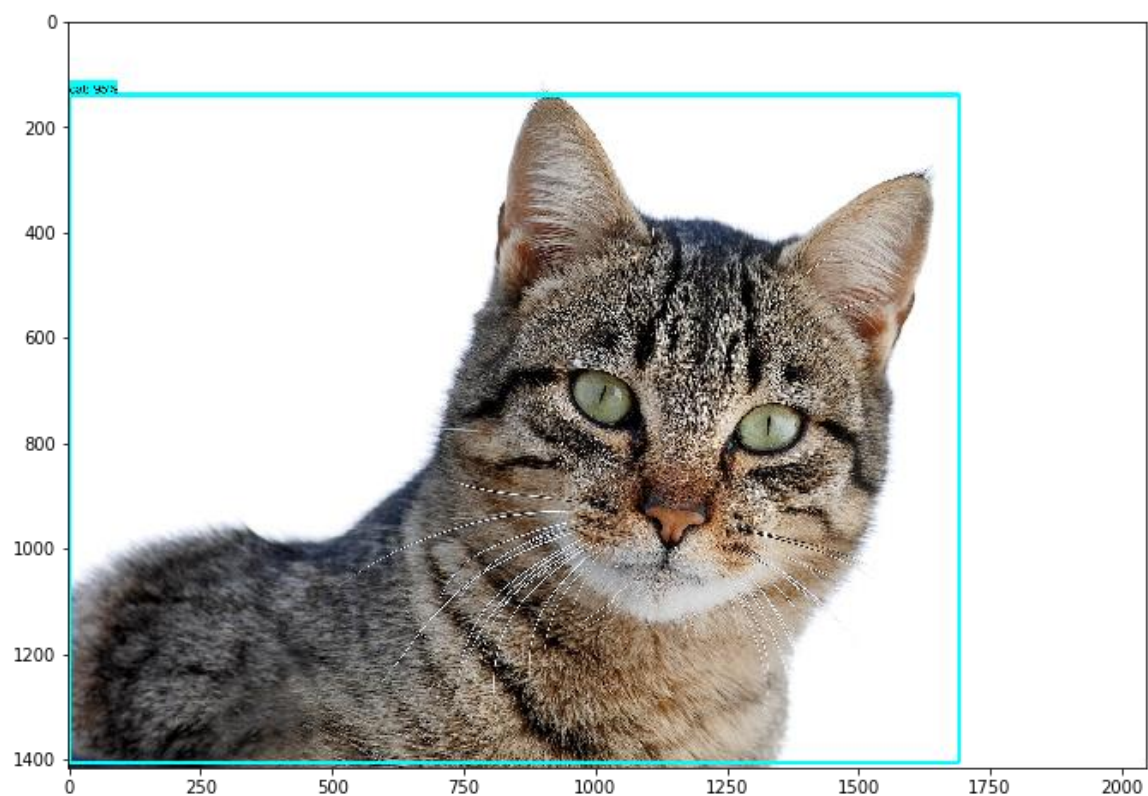


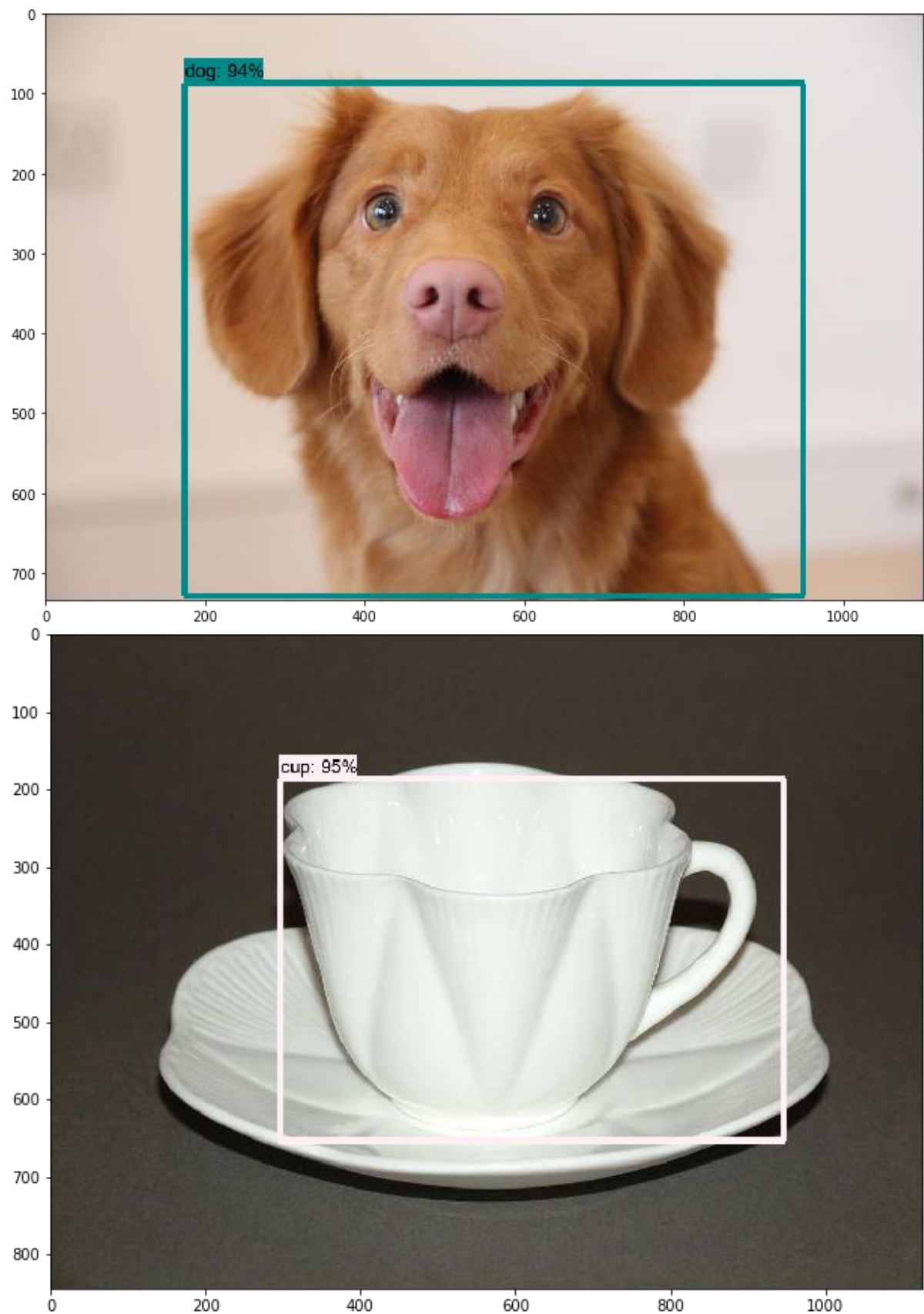
Slika 15: (Zatat, 2016)

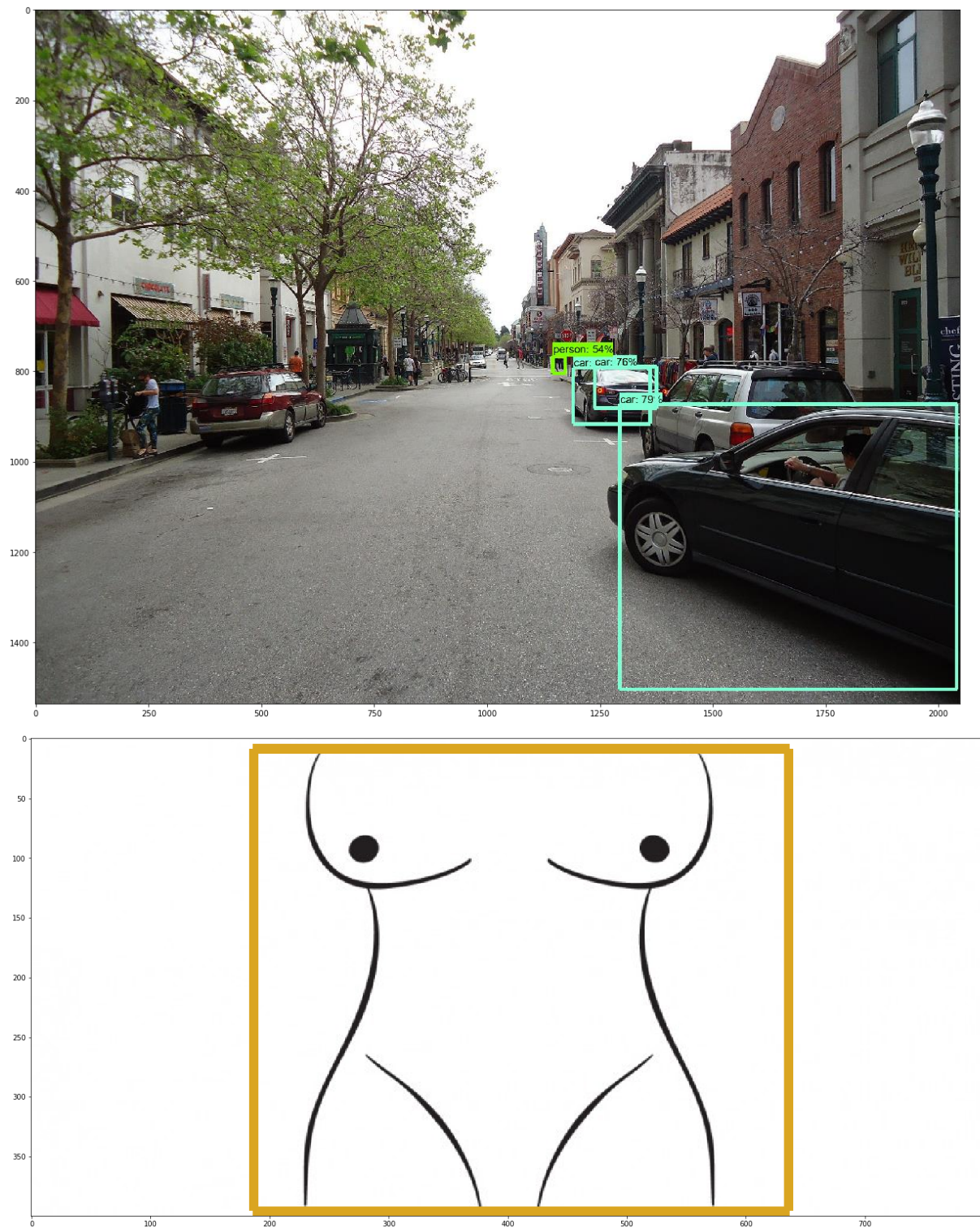


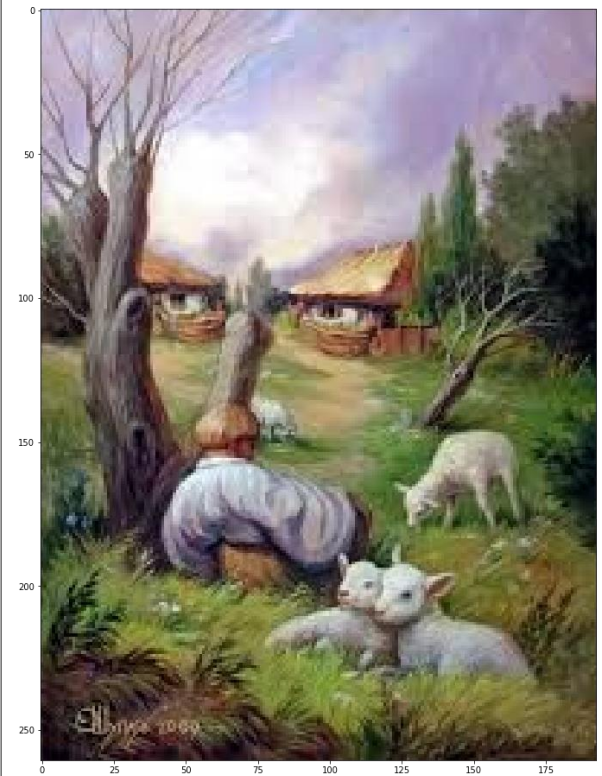
9.2 Priloga 2 – rezultati računalniške primerjave

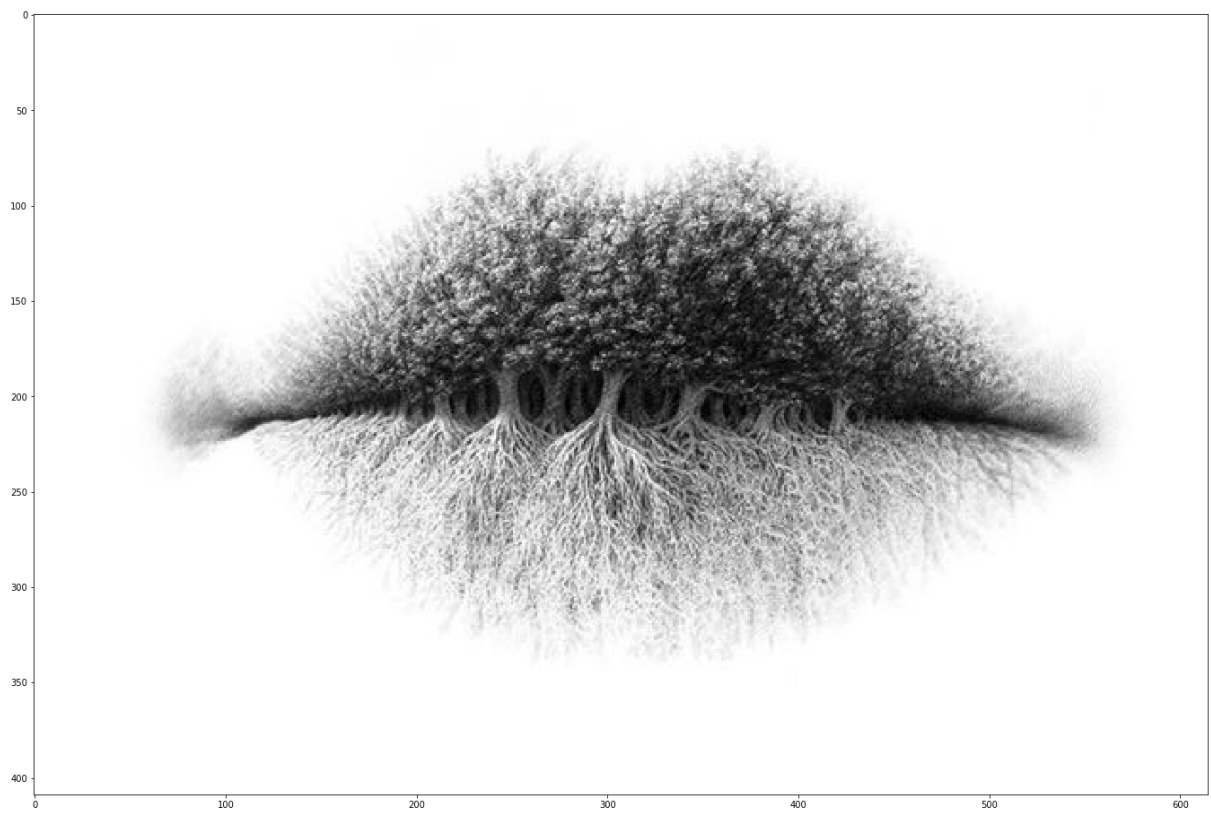
9.2.1 Slike 1/12



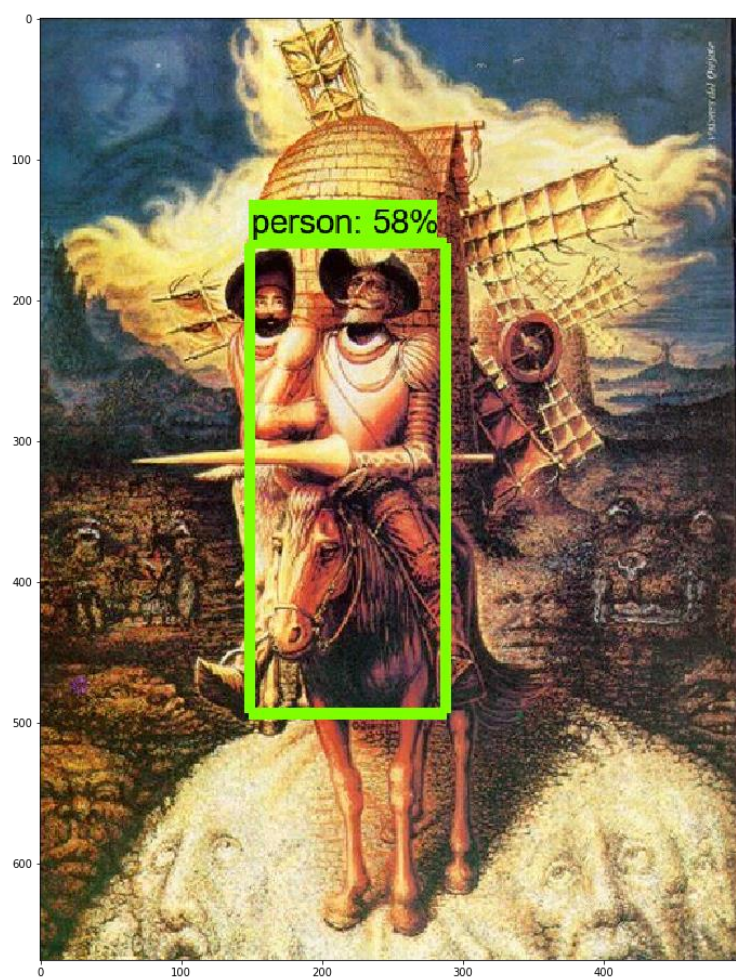
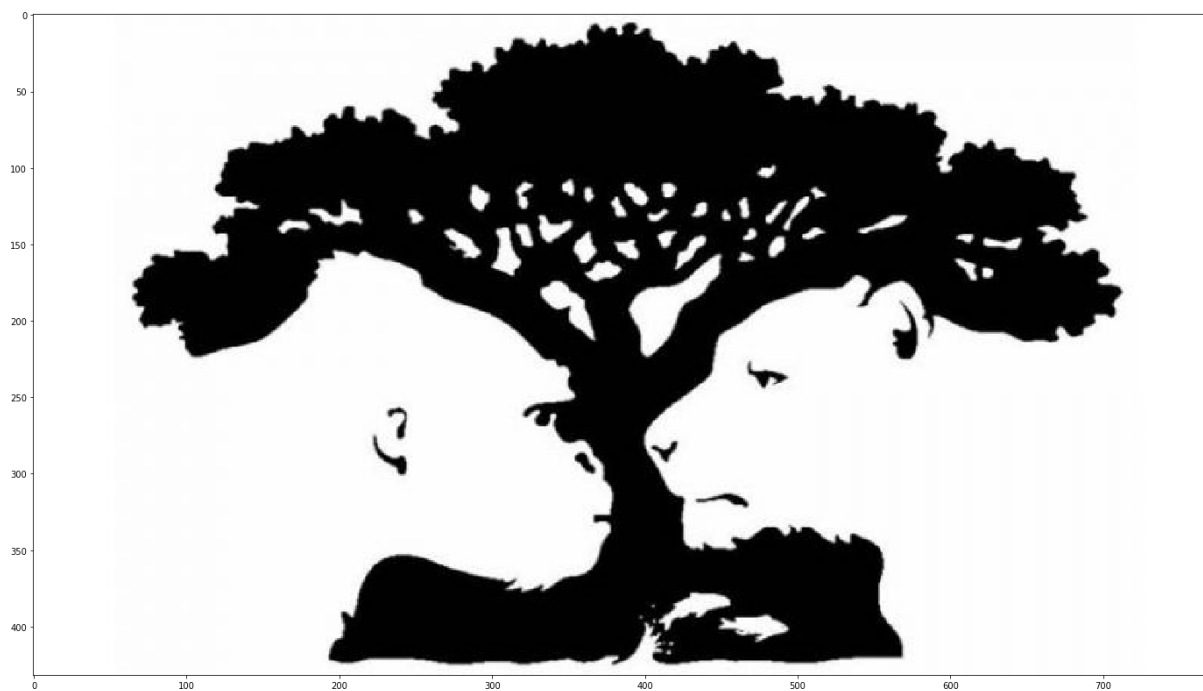




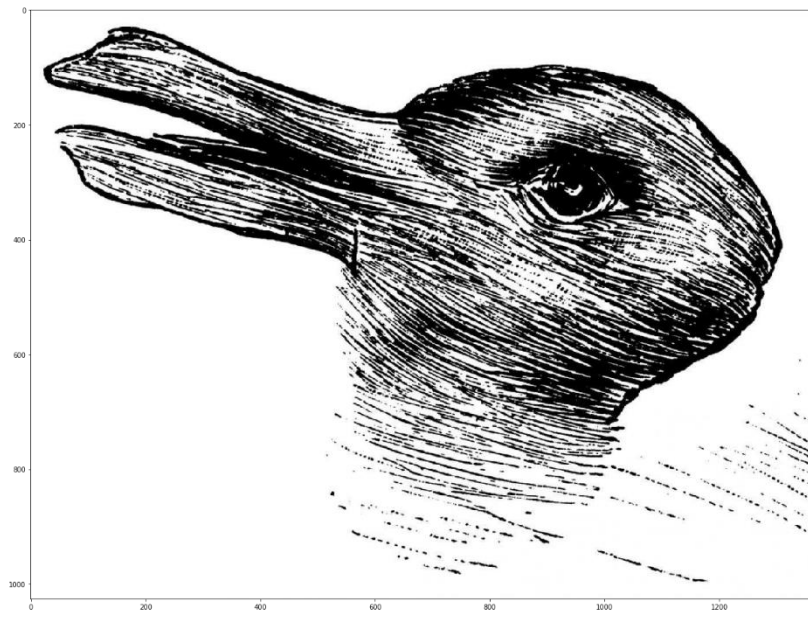












9.3 Priloga 3 – tabela rezultatov intervjuja

Row Labels	Count of ID		
slika1	69	slika11	69
cat	69	animals	5
slika2	69	fish	1
dog	69	lion	2
slika3	69	monkey	3
coffee	1	monkey and lion	1
mug	68	monkey and lion	9
slika4	69	monkey lion	9
car	6	monkey, tiger, tree	1
city	14	Panther and chimp	1
city	1	tree	37
new york	1	slika12	69
road	6	don quixote	16
street	37	einstein	2
street	2	face	7
street with cars	1	horse	8
town	1	human	2
slika5	69	knight on a horse	1
2 circles	1	knights	1
2 people dancng	1	man	7
body	25	man on a horse	1
body	37	man riding	2
N.G	1	man who rides	2
two friends waving at eachother	1	old man	3
woman	3	oldman	1
slika6	69	person	1
chair	69	rider	1
slika7	69	sancho pansa	2
countryside	4	towo men riding	1
face	26	two horsemen	1
face	1	windmill	10
farm	6	slika13	69
forest	1	caretaker	3
head	1	dog	6
man	11	dog	12
nature	1	dog feeding	2
nature	1	dog food	1
person	1	dog	1
sheep	10	doog feeding	1
tree	3	feeding	13
village	2	feeding	2
zeus	1	feedng	1
slika8	69	food	2
forest	18	friendship	1
forest	22	human	1
lips	26	love	1
roots	1	man	3
tree	2	man and a dog	13
slika9	69	man and a dog	1
airplane	69	man feeding dog	4
slika10	69	man feednig dog	1
car	67	slika14	69
mclaren	1	Hamburger	67
porche	1	Hamburger	2
		slika15	69
		bird	33
		bunny	16
		duck	18
		duck and bunny	2

Bibliografija

- Abrozzi. (2020). *What you see in this picture shows how sex-obsessed you are. What is it?* Pridobljeno 3. 1 2020 iz Abrozzi: <https://abrozzi.com/what-you-see-in-this-picture-shows-how-sex-obsessed-you-are-what-is-it/>
- Agencija za raziskovanje RS. (2009). *termania.net*. Pridobljeno 3. 1 2020 iz strukturirani intervju: <https://www.termania.net/slovarji/terminoloski-slovar-vzgoje-in-izobrazevanja/3474672/strukturirani-intervju>
- Anderton, H. (2020). *see-first-image-identifies-hidden-self*. Pridobljeno 3. 1 2020 iz Peace QUarters: <https://www.peacequarters.com/see-first-image-identifies-hidden-self/>
- COCO. (2019). *Common objects in context*. Pridobljeno 3. 1 2020 iz <http://cocodataset.org/#home>
- Cohut, M. (26. 8 2018). *Dogs: Our best friends in sickness and in health*. Pridobljeno 3. 1 2020 iz <https://www.medicalnewstoday.com/articles/322868.php>
- Dagorov, C. (2020). *mirror*. Pridobljeno 3. 1 2020 iz <https://www.mirror.co.uk/news/weird-news/what-you-see-photo-answer-16892307>
- Elfouly, S. (16. 7 2019). *Towards data science*. Pridobljeno 3. 1 2020 iz Part 1: R-CNN (Object Detection): <https://towardsdatascience.com/r-cnn-3a9beddfd55a>
- Google. (2019). *About Tensorflow*. Pridobljeno 3. 1 2020 iz Tensorflow: <https://www.tensorflow.org/about>
- Keras. (2019). *Home*. Pridobljeno 3. 1 2020 iz Keras: <https://keras.io/#why-this-name-keras>
- Khandelwal, R. (7. 12 2019). *Towards data science*. Pridobljeno iz COCO and Pascal VOC data format for Object detection: <https://towardsdatascience.com/coco-data-format-for-object-detection-a4c5eaf518c5>
- Khandelwal, R. (30. 11 2019). *Towards data science*. Pridobljeno 3. 1 2020 iz SSD : Single Shot Detector for object detection using MultiBox: <https://towardsdatascience.com/ssd-single-shot-detector-for-object-detection-using-multibox-1818603644ca>
- Narayan, T. (2019). *Object Detection : A Comparison of performance of Deep learning Models on Edge Using Intel Movidius Neural Compute Stick and Raspberry PI3*. Pridobljeno 3. 1 2020 iz Medium:

<https://medium.com/intel-student-ambassadors/object-detection-a-comparison-of-performance-of-deep-learning-models-on-edge-using-intel-f66eb7f45b17>

O'Reilly Media, Inc. (2020). *COCO object detection challenge*. Pridobljeno 3. 1 2020 iz O'Reilly: <https://www.oreilly.com/library/view/deep-learning-for/9781788295628/67acf64b-22e6-41cc-ac54-41331610e93b.xhtml>

Ocampo, O. (2020). *Face illusions*. Pridobljeno 3. 1 2020 iz Pinterest: <https://www.pinterest.com/pin/492229434242507039/?lp=true>

Pant, A. (7. 1 2019). *Introduction to Machine Learning for Beginners*. Pridobljeno 3. 1 2020 iz Towards data science: <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>

Petside. (14. 11 2019). *How To Choose The Right Type Of Food For Your Dog*. Pridobljeno 3. 1 2020 iz Petside: <https://www.petside.com/choosing-dog-food/>

Pinterest. (2020). *Illusion*. Pridobljeno 3. 1 2020 iz Pinterest: <https://www.pinterest.it/pin/555561304014875281/>

Redmon, J. (9. 5 2016). Pridobljeno 3. 1 2020 iz You Only Look Once: <https://arxiv.org/pdf/1506.02640v5.pdf>

Rugg, M. D. (2003). Human recognition memory: a cognitive neuroscience perspective. *Trends in cognitive sciences*, 7(7), 313-319.

Samuel, A. (7. 1 2019). *Introduction to Machine Learning for Beginners*. Pridobljeno 3. 1 2020 iz towards data science: <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>

SharpBrains. (2015). *Trend: Mobile apps to improve vision via perceptual learning*. Pridobljeno 3. 1 2020 iz SharpBrains: <https://sharpbrains.com/blog/2015/01/26/trend-mobile-apps-to-improve-vision-via-perceptual-learning/>

Sorokina, K. (2017). *Image Classification with Convolutional Neural Networks*. Pridobljeno 3. 1 2020 iz Medium: <https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8>

- Tyley, J. (23. 9 2015). *THE SCIENCE OF VISION: HOW DO OUR EYES SEE?* Pridobljeno 1. 3 2020 iz independent: <https://www.independent.co.uk/life-style/health-and-families/features/the-science-of-vision-how-do-our-eyes-see-10513902.html>
- Wei Liu, D. A.-Y. (8. 12 2015). *Cornell University*. Pridobljeno 3. 1 2020 iz SSD: Single Shot MultiBox Detector: <https://arxiv.org/abs/1512.02325>
- Wikimedia Commons, the free media repository. (1. 1 2016). *File:Cat March 2010-1.jpg*. Pridobljeno 3. 1 2020 iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:Cat_March_2010-1.jpg&oldid=183411293.
- Wikimedia Commons, the free media repository. (25. 11 2016). *File:Cup and saucer Dainty White design by Rowland Morris 1896.jpg*. Pridobljeno iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:Cup_and_saucer_Dainty_White_design_by_Rowland_Morris_1896.jpg&oldid=219498919.
- Wikimedia Commons, the free media repository. (15. 12 2017). *File:Boeing 787-8 maiden flight takeoff.jpg*. Pridobljeno iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:Boeing_787-8_maiden_flight_takeoff.jpg&oldid=272161639.
- Wikimedia Commons, the free media repository. (12. 11 2018). *File:Air chair-edit-1-2018.jpg*. Pridobljeno 3. 1 2020 iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:Air_chair-edit-1-2018.jpg&oldid=327498792.
- Wikimedia Commons, the free media repository. (15. 3 2018). *File:Wendy's Double Bacon Deluxe hamburger.jpg*. Pridobljeno 3. 1 2020 iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:Wendy%27s_Double_Bacon_Deluxe_hamburger.jpg&oldid=292291867.
- Wikimedia Commons, the free media repository. (21. 3 2019). *File:California Santa Cruz street with cars and shops.JPG*. Pridobljeno iz Wikimedia Commons, the free media repository: https://commons.wikimedia.org/w/index.php?title=File:California_Santa_Cruz_street_with_cars_and_shops.JPG&oldid=343405110.

Wikimedia Commons, the free media repository. (18. 8 2019). *File:Porsche 911 GT3 Touring, IAA 2017, Frankfurt (1Y7A2766).jpg*. Pridobljeno 3. 1 2020 iz Wikimedia Commons, the free media repository:

[https://commons.wikimedia.org/w/index.php?title=File:Porsche_911_GT3_Touring,_IAA_2017,_Frankfurt_\(1Y7A2766\).jpg&oldid=362459317](https://commons.wikimedia.org/w/index.php?title=File:Porsche_911_GT3_Touring,_IAA_2017,_Frankfurt_(1Y7A2766).jpg&oldid=362459317).

Zatat, N. (2016). *What do you see in this picture? The answer says a lot about your personality*. Pridobljeno 3. 1 2020 iz Indy100: <https://www.indy100.com/article/what-do-you-see-in-this-picture-the-answer-says-a-lot-about-your-personality-7316291>