

**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U  
OSIJEKU  
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA  
I INFORMACIJSKIH TEHNOLOGIJA OSIJEK**

**Sveučilišni studij**

**PRIMJENA NEURONSKIH MREŽA U  
KLASIFICIRANJU LEGO KOČAKA**

**Diplomski rad**

**Anto Tufeković**

**Osijek, 2021.**

## SADRŽAJ

|  |    |
|--|----|
| 1. UVOD .....  | 1  |
| 1.1. Zadatak diplomskog rada .....                           | 2  |
| 2. PREGLED PODRUČJA RADA.....                                | 6  |
| 3. SUSTAV ZA KLASIFICIRANJE LEGO KOCAKA .....                | 10 |
| 3.1. Implementacija sustava za generiranje podataka .....    | 10 |
| 3.1.1 Generiranje podataka .....                             | 10 |
| 3.2. Usporedba postojećih skupova podataka .....             | 12 |
| 3.2.1 Generirani skup podataka i ostvareni rezultati.....    | 12 |
| 3.3. Implementacija ANN za klasifikaciju .....               | 15 |
| 3.3.1. Opis slojeva i strukture .....                        | 17 |
| 3.4. Usporedba s drugim umjetnim neuronskim mrežama.....     | 25 |
| 4. DEMONSTRACIJA I ISPITIVANJE FUNKCIONALNOSTI SUSTAVA ..... | 27 |
| 4.1. Demonstriranje mogućnosti sustava.....                  | 27 |
| 4.2. Ispitivanje funkcionalnosti sustava .....               | 29 |
| 4.3. Analiza rezultata.....                                  | 31 |
| 5. ZAKLJUČAK .....   | 35 |
| PRIZNANJA.....   | 37 |
| LITERATURA .....   | 38 |
| SAŽETAK .....  | 41 |
| ABSTRACT.....  | 42 |
| ŽIVOTOPIS.....   | 43 |
| PRILOG A: Popis korištenih modela i njihovi prikazi .....    | 44 |

## 1. UVOD

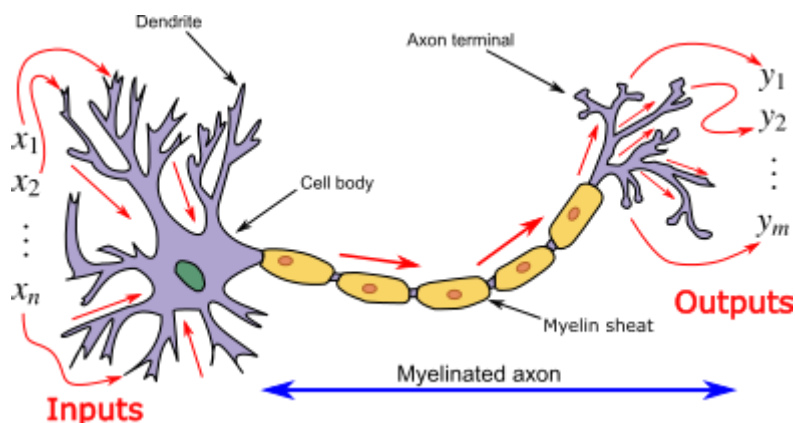
Često se moraju odrediti točni detalji o nekom objektu ili stvorenju, primjerice psa ili mačke gdje vlasnik ne zna dovoljno o pasminama da utvrdi točno kojoj pripada. Za taj problem mogu poslužiti umjetne neuronske mreže koje su trenirane nad više tisuća različitih slika svih vrsta pasmina domaćih životinja tako da za neku ulaznu sliku (npr. koju vlasnik uslika) preda izlaznu oznaku koja govori o vjerojatnostima kojim pasminama životinja na slici pripada. Isti princip prepoznavanja objekta sa slike se može primijeniti na razne stvari u svijetu koje sadrže uočljive razlike kao što su modeli automobila po markama, obitelji vatrenih oružja, vrsta sorte voća (npr. jabuka i krušaka), vrste glava šarafa te vrste i modeli igrački, koji su objekt promatranja u ovom radu.

Jedna od situacija gdje ova tehnologija može pomoći je prilikom dizajniranja npr. mobilne aplikacije za slijepe ili slabovidne osobe. Osobe s poteškoćama vida i osjetila dodira (svi formalni lijekovi sadrže ime u Brailleovom pismu) mogu iskoristiti ovakve tehnologije za npr. prepoznavanje kutije lijekova po njenim karakteristikama (veličina kutije, boja kutije, šare po kutiji, prepoznavanje teksta na kutiji) da odaberu točne lijekove u pripadajućem trenutku.

Cilj ovog diplomskog rada je istražiti izradu umjetne neuronske mreže za prepoznavanje i klasificiranje LEGO kocaka na slikama, na takav način da svakoj ulaznoj slici da izlaznu oznaku koja govori kojoj klasi pripada (klasifikacija s više klasa, engl. *multiclass classification*). Potrebno je napraviti pregled problema na koje se naiđe tijekom treniranja takve umjetne neuronske mreže i postaviti ili izvesti rješenja problema.

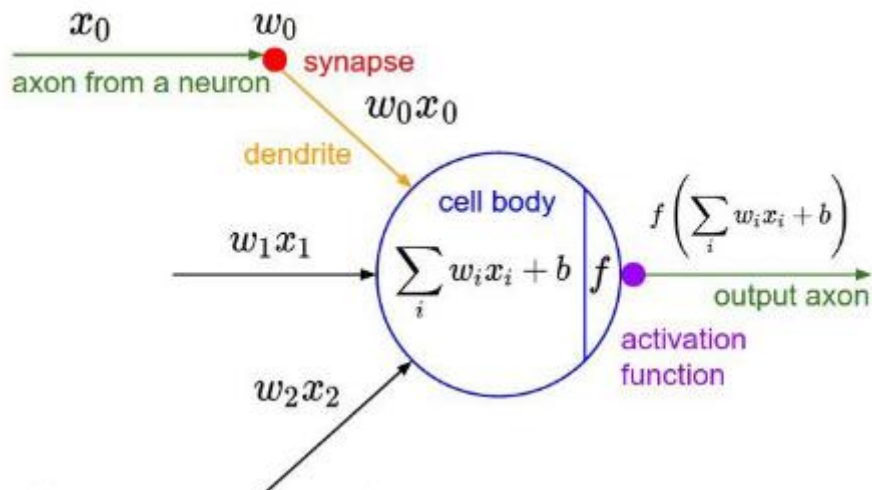
## 1.1. Zadatak diplomskog rada

U ovom radu će se koristiti **duboke umjetne neuronske mreže** (engl. *deep artificial neural networks*) za potrebe razvoja algoritma za klasifikaciju. Umjetne neuronske mreže su vrsta algoritma koji su bazirani nad neuronskim mrežama životinja i njihova osnovna građevinska jedinica (neuron) je vrlo slična biološkom neuronu prikazanom na slici 1.1.



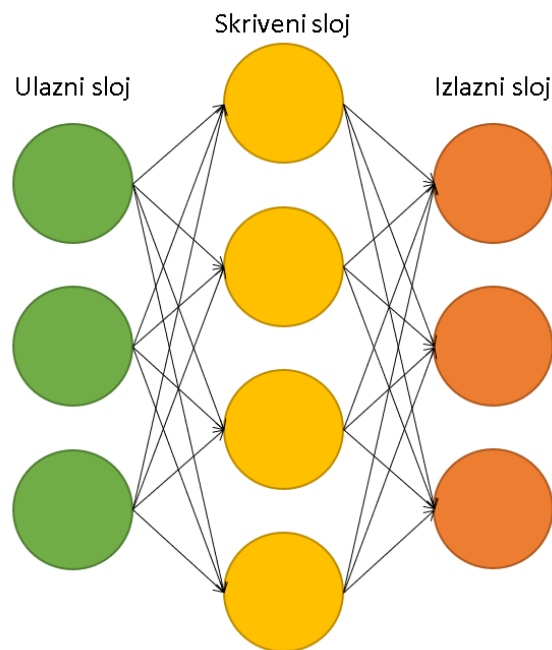
**Slika 1.1.** Uzor neurona je prava biološka živčana stanica [1]

Na slici 1.2. se vidi da svaki umjetni neuron po uzoru biološkog neurona ima ulaz, prijenos i izlaz. Ulazi ( $x_1, x_2$ ) na dendritima primaju signale iz prethodnih neurona u mreži. Svaki neuron sadrži u sebi prijenosnu funkciju kojom djeluje nad ulazima i onda na izlazima ( $y_1, y_2$ ) propagira dalje signale. Prethodno spomenuta funkcija stanice neurona određuje kakav će signal biti na svakom izlazu ovisno o ulazu, funkcija sadrži u sebi težine koje se pri treningu mijenjaju ovisno o stimulusu i onda odlazi u aktivacijsku funkciju. Najjednostavnije objašnjenje je da svaki ulaz sadrži svoju težinu koja množi, tj. pojačava ili oslabljuje signal tog danog ulaza.



**Slika 1.2.** Matematički model neurona [2]

Osnovni matematički model neurona je da svaki ulaz ima svoju težinu [3], težine se postavljaju tijekom treninga ovisno o važnosti ulaza. Na kraju neuron izvršava neku funkciju (engl. *activation function*) nad sumom svih ulaza množeni svojim težinama i to predstavlja kao izlaz svim sljedećim spojenim neuronima.



**Slika 1.3.** Primjer neuronske mreže s jednim skrivenim slojem

Kao što je prethodno spomenuto, neuroni su često poslagani u slojevima zbog jednostavnosti kreiranja modela, no to ne znači da signal putuje jednom kroz model i završava odmah na izlazu, nego u modelu se mogu nalaziti razne funkcionalnosti između slojeva koje pomažu

modelu postići bolje rezultate. Glavni cilj tih slojeva, spojeva i raznih struktura je izbjegavanje mrtvih neurona (neuroni čije su težine ulaza postavljene blizu granice aktivacijske funkcije, npr. vrijednost jako blizu nuli za *ReLU*).

**Klasifikacija** je problem iz područja statistike. U problemu se pokušava doći do strukturiranog rješenja gdje nekoj promatranoj jedinici damo jednu oznaku iz skupa oznaka. Jedan jednostavan primjer bi bila klasifikacija elektronske pošte u ulazni sandučić ili u neželjenu poštu (engl. *spam*). Takvi ulazi moraju u sebi sadržavati neku informaciju (engl. *feature*) koja algoritmu pomaže odrediti kojoj od klasa ulazni podatak pripada. Takve informacije mogu biti same po sebi kategorične (npr. osoba može imati krv tipa „A“, „AB“, „A+“, itd.), mogu biti zasnovane na brojevima (npr. slike imaju piksele koje sadrže tri kanala za boje, najčešće s vrijednostima od 0 do 255) ili jednostavno uspoređivati dolazeće podatke s prijašnje treniranim podacima tako da uspoređuje udaljenost ili sličnost novih i prethodnih obrađenih podataka.

Algoritam koji obavlja klasifikaciju se naziva klasifikator. Ako postoje samo dvije klase za klasificirati onda se metoda zove „binarna klasifikacija“ (engl. *binary classification*). Ako imamo više klasa metoda se zove „klasifikacija s više klasa“ (engl. *multiclass classification*), a ako imamo metodu gdje neki izlaz može imati više izlaznih klasa pridruženo onda se radi o „klasifikaciji s više oznaka“ (engl. *multilabel classification*).

Za potrebe treniranja neuronskih mreža potrebna je velika količina prethodno ručno označenih (engl. *annotated*) podataka. Označavanje podataka je vremenski zahtjevan zadatak i česti su slučajevi da **ne postoji dovoljno dostupnih podataka nad kojima se može trenirati**. Razni su razlozi kao npr. poštivanje privatnosti u medicini (npr. rendgenske, mamografske, ultrazvučne slike i slično) zbog čega gubimo pristup vrijednim podacima za treniranje, a potrebna je velika količina da se može trenirati kvalitetna mreža, pogotovo ako je duboka mreža. Dodatan problem je prethodno navedeno označavanje podataka, nije moguće komplicirane medicinske snimke predati nekoj osobi za označavanje kad je potrebno znanje stručnjaka za pravilno označavanje npr. tumora na ultrazvuku. Što je dublja umjetna neuronska mreža to je više raznolikosti u trening podacima potrebna da se ona kvalitetno trenira bez problema pretreniranja (engl. *overfit*) i podtreniranja (engl. *underfit*).

Kako bi se ublažio problem manjka podataka, skup podataka se može na razne načine proširiti. Za tu svrhu postoji proširenje skupa podataka (engl. *Data augmentation*). To je svaka vrsta tehnike proširenja postojećeg skupa podataka tako da se uvede nasumično transformirana kopija nekog originala iz skupa ili se generira sintetički podatak od originalnih[4] uzoraka. Za

kreiranje nasumične kopije originala često se koriste jednostavne funkcije kao što su affine transformacije (rotacija, skaliranje, translacija, refleksija, smicanje), transformacije nad histogramom (npr. ujednačenje) i mijenjanje svojstva slike kao što su kontrast, svjetlina ili transformacije kao povećanje oštine, pomućenje slike i izjednačavanje svjetline.

Razlog za proširenjem skupa podataka je poboljšanje generalizacijskog svojstva mreže koje se ostvaruje povećanjem dostupnog broja uzoraka podataka. Ako se kompleksna mreža trenira nad malim brojem podataka za normalan broj epoha, dogodit će se podtreniranje (engl. *underfitting*) mreže jer mreža nije dovoljno trenirana, tj. nije prošla kroz dovoljno primjeraka da se namjestite težine neurona da uspješno prepozna objekt na slici. Ako se trenira nad malim brojem podataka za više epoha mreža će početi „pamtiti“ ulazne podatke, njene težine se namještaju na takav način da ima što manju moguću grešku (engl. *loss*) i dogodit će se pretreniranje (engl. *overfitting*). Pretreniranje smeta generalizaciji, jer pri treniranju će trenirana mreža imati visoku točnost, a pri validiranju nisku točnost.

Ovaj rad je usredotočen na klasificiranje LEGO kocaka nad slikama. Svaki model LEGO kocke se predstavlja kao klasa (npr. 3001 model i 3003 model su sami svoje klase), tip je glavno područje modela (npr. 3001 i 3003 modeli su tipa ravna kocka, 3020 i 3022 su tipa ravna ploča, 6632 i 32140 su tehničke kocke, itd.) i kad bi se ciljali svi modeli svih setova ikad izdanih, postojalo bi oko 4000 i naviše klasa. Takav visok broj modela bi iziskivao značajne računalne resurse. Za potrebe ovog rada i za potrebe iskazivanja mogućih problema perspektive uzima se u obzir 25 odabranih modela kao klase ovog rada. Svi modeli lego kocaka korišteni za ovaj rad i njihovi grafički prikazi nalaze se u prilogu A

Za potrebe rada odabrane su LEGO kocke jer one sadržavaju jednostavne značajke (veličina, jasni rubovi, dugmadi) nad kojima se duboke neuronske mreže mogu trenirati. Treniranje se provodi nad sintetički generiranim podacima, zbog korištenja takvih podataka nailazi se na problem razlike između realnih i sintetičkih podataka [5] (engl. *domain gap*).

Rad je podijeljen u 3 glavna poglavlja i zaključak. Prvo glavno poglavlje se bavi opisom trenutnog stanja aktualnih radova (engl. *state of the art*). Opisuje se postojeća rješenja za LEGO klasifikaciju i dodatni radovi znanstvene zajednice koji su vezani za ovaj rad. Drugo glavno poglavlje opisuje generiranje sintetičkih podataka za potrebe treniranja umjetnih neuronskih mreža i bavi se s implementacijom dubokih umjetnih neuronskih mreža koristeći pripadajuće alate (Python, Keras, itd.). U zaključku se nalaze izlazne riječi autora s mišljenjima o radu, interpretacija rezultata, pregled doprinosa i analiza prednosti i mana korištenog pristupa.

## 2. PREGLED PODRUČJA RADA

Klasificiranje slika su ljudi izvodili već duže vrijeme po [6], AlexNet [7] je jedan od brojnih primjera ranijih neuronskih mreža. Ona je 2012. godine u ImageNet natjecanju ILSVRC-2012 postigla prvo mjesto. AlexNet je postigla 58% manje grešaka (engl. *top-5 error rate*) naspram sljedećeg najboljeg (sa 26.2% na 15.3%), najviše jer su koristili duboke neuronske mreže koje su trenirane koristeći grafičke kartice zbog velikih zahtjeva za računalnim resursima. Ovo je privuklo globalnu pozornost prema koristi područja strojnog učenja i umjetne inteligencije te ubrzo je krenulo napredovanje tih tehnologija. Danas se rezultati tehnologija dubokih konvolucijskih neuronskih mreža nalaze u raznim poljima svakodnevnice. Kamere koriste tehnologije prepoznavanja lica i aktivnog stabiliziranja videosnimke, doktori s modelima mogu lakše uočiti bolesti na raznim vrstama snimaka, od EKG podataka do rendgenskih slika i moguće je lakše prevesti pokrete tijela u pokrete objekata virtualne stvarnosti uz pomoć neuronskih mreža (npr. za koristi u arhitekturi [8]).

Joseph Redmon iz Sveučilišta Washington je koautor rada „YOLO9000: Better, Faster, Stronger“ [9] gdje opisuje svoju mrežu YOLO9000, koja u stvarnom vremenu može detektirati i klasificirati objekte s preko 9000 kategorija klasa. U radu opisuje brojna poboljšanja nad originalnim YOLO modelom sa YOLOv2 kojeg iskušava na skupovima podataka PASCAL VOC [10] i COCO [11]. Nad PASCAL VOC skupom podataka model radi na 67 okvira po sekundi i dobiva performanse od oko 76.8 mAP, dok nad COCO skupom podataka radi na 40 okvira po sekundi i dobiva performanse od oko 78.6 mAP.

Modificirana VGG16 mreža Jacquesa Mattheija [12] je mreža za klasifikaciju tipa modela LEGO kocke na slici (je li je dana slika modela kocke, ograde, *Technic* bloka, vegetacije, ono daje tip, ne direktno model.). Na 60000 primjera podatkovnog skupa za treniranje ima točnost od oko 95%. Točnost je veća za klase koje se često pojavljuju unutar skupa podataka, dok je manja za klase koje su slabije zastupljene kao što su npr. rotacijska ploča (engl. *turntable*). Razlika u točnosti među mrežama dolazi od toga da Mattheijova mreža ne mora točan model odrediti, makar mu se da slika kocke sa strane ili od gore, ono će samo razmišljati kojem skupu pripada i taj tip LEGO kocke sortirati ovisno o rezultatu predviđanja u odgovarajuću košaru (košare za kocke, tanke ploče, široke ploče, vegetaciju, figurice, zastave, itd.) koristeći mehanički stroj s pokretnom trakom. Stroj promatra unesene kocke na pokretnoj traci, kocke se promatraju kamerom koja je spojena na računalo koje putem duboke konvolucijske



neuronske mreže predviđa klasu promatrane kocke. Ovisno o predviđenoj klasi, kocke se sortiraju u određenu košaru koristeći mlaznice komprimirani zrak.

Još jedan primjer stroja je stroj za sortiranje kojeg je izradio Daniel West [13] iz inspiracije prethodno spomenutog Mattheijovog stroja. West spominje da je njegov stroj „sljedeći korak u evoluciji“, po tome što ima sposobnost prepoznavanja LEGO kocaka po modelu. To znanje se koristi u stroju za sortiranje tako da se modeli grupiraju po tipu objekta i spremaju se (slično kao Jacquesov stroj) u odgovarajuću košaru ovisno o predviđenoj klasi (kocke u kocke, grede s gredama, komadi figurica s ostalim komadima, itd.). Uspjeh stroja leži u načinu kako je koncipiran. Originalno je treniran nad sintetičkim podacima i kad mu se daju stvarne slike iz pravog svijeta, ima lošu točnost predviđanja. Vidjevši da korištenje unaprjeđenja skupa podataka nema zadovoljavajuće rezultate, West je prestao pokušavati generirati što sličniju sliku realnom i krenuo je generirati slike koje se više koncentriraju na specifične detalje modela u pokušaju smanjivanja razlike domena stvarnih i sintetičkih slika. S povećanjem točnosti na kraju je West počeo koristiti bilježenje oznake uz pomoć njegove mreže, mreža mu je ponudila što ono misli da je najbliže predviđanje pa West mora odabrati koji je to točno model i na tom odabiru se mreža uči (nadzirano učenje). West tvrdi da je oko 100,000 slika uspješno označio na ovaj način. Jedini problem je što West nije objavio nikakve podatke, metrike ili označene slike u javnosti i zbog toga se ne može pravilno uspoređivati s dobivenim mrežama, iako postoje razni artikli i par videozapisa [14] koje je West postavio iz čega se vidi efektivnost stroja.

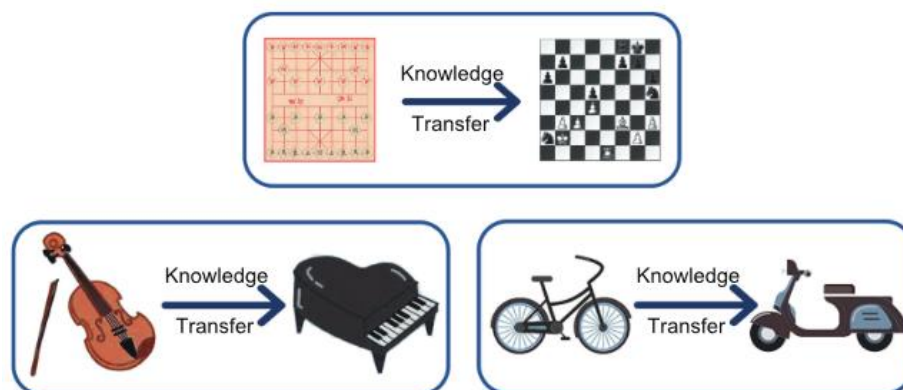
Rad kojeg je napisala Agnieszka Mikołajczyk u Gdansk Sveučilištu tehnologije [4] opisuje razne metode i primjene proširenja skupa podataka za poboljšanje rezultata mreža za duboko učenje. Rad se temelji na dvije vrste proširenja skupa podataka: iskrivljavanje podataka i korištenje drugih mreža za generiranje podataka (engl. *oversampling*). Iskrivljavanje podataka se osniva na korištenju transformacija kao što su afine transformacije: translacija, rotacija, refleksija i smicanje s nasumičnom razinom učinka (npr. translacija se može izvesti samo u rasponu  $\pm 20\%$ , smicanje samo  $\pm 15^\circ$ , itd.) te korištenje transformacija nad kanalima boje u slici kao što su izjednačavanje histograma, mijenjanje kontrasta i mijenjanje „topline“ bijele boje. Još neke dodatne transformacije koje se mogu izvesti su izoštravanje (engl. *sharpen*) i zamućenje (engl. *blur*) slike. *Oversampling* se osniva na generiranju sasvim novih podataka učeći uzorke od postojećih ulaznih podataka. Jedna od tehnika *oversamplinga* se naziva GAN (engl. *Generative Adversarial Network*). GAN i općenito *oversampling* nisu optimalni pristupi za mreže koje trebaju prepoznavati značajke slične LEGO kockama jer GAN-ovi ne mogu

pravilno naučiti generirati umjetne podatke modela koji su strogo povezani sa svojim osima (npr. pravokutnike, trokute, kocke) ili strogom pozicijom ili veličinom teksture na modelu (u našem slučaju pozicija dugmadi LEGO kocke mora biti na vrhu kocke i da količina dugmadi odgovara veličini LEGO kocke). Iz ovog rada se koristi generalno znanje da čak s malim učinkom proširenja skupa podataka se poboljšavaju rezultati generalizacije dubokih mreža i dodaje motivaciju za istraživanje optimalnog skupa parametara s kojim se mogu postići bolji rezultati.

Dodatan rad koji opisuje proširenje skupa podataka su napisali korisnici časopisa *Journal of Big Data* Connor Shorten i Taghi M. Khoshgoftaar [15]. Ono slično prethodno navedenom radu opisuje probleme nedostataka ulaznih podataka za treniranje dubokih mreža (npr. u polju medicine) i istražuje moguća rješenja. Isto se dijeli na iskrivljavanje podataka i *oversampling*. U kontekstu iskrivljavanja podataka veliku pažnju rad obraća na sigurnost primjene neke transformacije nad ulaznim podacima (sigurnost u smislu da ulazni podatak ne izgubi svoju konekciju nad oznakom koju nosi, npr. u MNIST [16] skupu podataka za rukom pisane brojeve, važno je ne primjenjivati refleksiju i držati rotaciju unutar prihvatljivih granica, inače se može dogoditi pobuna oko nekih brojeva kao 6 i 9 ako se previše rotira [15, str. 8]). Još jedan problem koji opisuje je kada su ulazni podatci savršeno centrirani, duboke mreže učene nad takvim podacima zahtijevaju da su novi podatci za predviđanje isto savršeno centrirani. Sličan problem se nalazi u ovom radu jer sve generirane slike su centrirane. Translacijom se taj problem do neke razine uklanja (npr. imat će i dalje poteškoće prepoznavati modele na samom rubu slike). Jedno od tehnika koju ovaj rad primjenjuje od spomenutog rada jest zamućivanje pozadina ulaznih slika, jedan od problema koji mogu nastati korištenjem ove duboke mreže jest slikanje, npr. kamerom mobitela, neke LEGO kocke. Sama LEGO kocka će ostati oštra ali pozadina neće biti u fokusu i bit će zamućena. U ovom radu se pozadinske slike lagano zamućuju prilikom učenja da bi se simulirao gubitak fokusa pozadine na kameri.

Rad Fuzhen Zhuang iz Instituta računalne tehnologije Peking [17] opisuje prijenosno učenje (engl. *transfer learning*) u znanstvenom polju dubokih mreža. Prijenosno učenje je koncept prenošenja korisnih informacija iz jedne domene učene duboke mreže u drugu domenu koja sadrži visok stupanj korelacije. Npr. iz mreže koja je učena prepoznavati oblike prenesemo težine na mrežu koja uči LEGO kocke. Zato što postoje naučeni identifikatori rubova učenje nove mreže će se ubrzati. Ovaj postupak se može izvoditi samo među mrežama sa sličnim ulaznim domenama (po slici 2.1., kao što su: vožnja bicikla se može prenijeti na vožnju motora,

sviranje instrumenta svoje obitelji, kao što je flauta, na sviranje instrumenta sličnog tipa, kao klarinet, itd.).



**Slika 2.1.** – Prikaz mogućnosti prijenosa znanja za potrebe prijenosnog učenja [17 str.1]

Rad opisuje prednosti prijenosnog učenja, gdje ako nova mreža ima manjak ulaznih podataka na kojima se može učiti da koristeći prijenosne mreže može pomoći ispuniti tu početnu prazninu treniranja, iako manjak ulaznih podataka i dalje smeta razvoju nove mreže jer mora razbiti učenje u zadnjim konvolucijskim blokovima prenošene mreže da bi bila efektivna. Jedan od problema prijenosnog učenja je događaj „negativnog prijenosa“ [18], gdje prethodno naučeno znanje mreže smeta učenju nove mreže. Ovo je rijedak događaj ali uvijek je moguće kada postoji manjak označenih podataka za učenje.

U ovom radu se koriste razni koncepti opisani u prethodno navedenim radovima, zbog manjka raznolikosti podataka (postoji malen broj skupova podataka LEGO kocaka, generirani skup je temeljen nad skupom od korisnika Joost Hazelzet sa stranice Kaggle [19]) se koristi proširenje skupa podataka koristeći tradicionalne metode (opisane u gore spomenuta dva rada kao afine transformacije) i koristi se prijenosno učenje za svrhe istraživanja tih mogućnosti kao najbolje rješenje u ovom radu.

### 3. SUSTAV ZA KLASIFICIRANJE LEGO KOCAKA

#### 3.1. Implementacija sustava za generiranje podataka

*Blender* [20] je program za 3D modeliranje i sadrži mogućnosti za izradu slika i animacija modela. Koristi se u ovom radu za generiranje slika sintetičkog skupa.

Cijeli projekt za ovaj rad se nalazi na sljedećoj *github* poveznici:

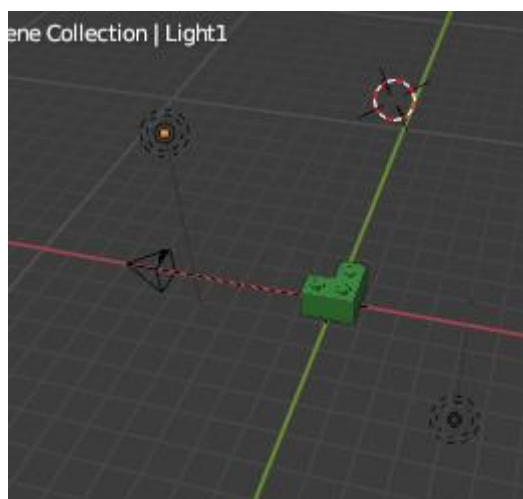
[https://github.com/ATufekovic/LEGO\\_classification](https://github.com/ATufekovic/LEGO_classification)

Projekt sadrži dva direktorija, *docs* i *py*. Unutar *docs* se nalaze svi materijali vezano za pisani rad i ovaj sam rad. Unutar *py* se nalazi više poddirektorija, svaki poddirektorij sadrži svoju *Jupyter* bilježnicu unutar koje se izvršavao *Python* kôd. Direktoriji s prefiksom „*test*“ su direktoriji koji sadrže neuronske mreže nad kojima se treniralo i sve funkcije vezane za njih. Direktorij „*image\_generation*“ sadrži *Blender* skriptu i *.blend* datoteku za koju je skripta namijenjena. Datoteka sadrži sve modele i okolinu za slikanje slika, dok skripta služi za automatizaciju generiranja slika. Direktorij „*image\_sourcing*“ sadrži *Jupyter* bilježnicu za generiranje slika za korištenje u ovom dokumentu. Ovo stoji tu za znatiželjne koji žele pregledati sve vezano za rješenje ovog rada.

Na *github* poveznici se ne nalazi dobiveni skup podataka niti dobivena mreža zbog ograničenja veličine. *Github* se ne bi trebao koristiti kao repozitorij skupova podataka i jako velikih datoteka.

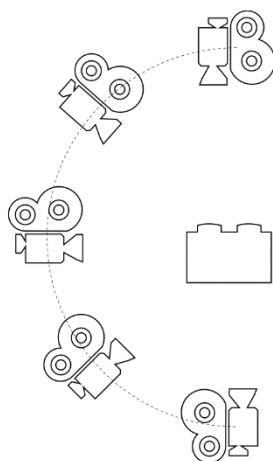
##### 3.1.1 Generiranje podataka

U *Blenderu* je potrebno učitati gore spomenutu datoteku scene, unutar nje se nalaze svi potrebni modeli ovog rada. Sve što korisnik skripte mora učiniti je prenijeti modele iz nevidljive skupine u vidljivu i obratno kako je poželjno. Sintetički podatci za ovaj rad su generirani u *Blenderu* koristeći skriptu koja je namještena da uslika zadani model iz raznih kutova. Generirane slike sadrže dva svjetlosna izvora u blizini sredine ali su za jednu prostornu jedinicu odmaknute svaka u svoju stranu (na slici 3.1. se vidi da su odmaknute od centra 4 prostorne jedinice, u stranu pomaknute za 1 prostornu jedinicu i visoko tj. duboko postavljeno 6 prostornih jedinica).



**Slika 3.1.** Isječak iz Blender prozora za prikaz lokacija objekta, svjetla i kamere

LEGO modeli se skaliraju tako da imaju točan odnos veličine međusobno (npr. vizualno pregledati da su na svim modelima dugmadi i zajedničke dužine iste i time garantirano iste skale, da su tehničke kocke iste dužine (npr. 2M, 3M), itd.). LEGO modeli su svi slikani s jednakom udaljenošću od kamere. LEGO modeli su nasumično obojani od par odabranih klasičnih LEGO boja, kao što su plava, crvena, siva, bijela, itd. Neuronske mreže moraju biti otporne na različite boje u slikama, iako jedno moguće rješenje problema je da se slike prije treniranja pretvore u slike sivih tonova.



**Slika 3.2.** Prikaz načina kretanja kamere u Blenderu

Kamera sliká model počevši od skroz gore (slika 3.2.) i u zadanim koracima se kreće sferično oko modela prema dolje. Kako se kamera kreće, tako ono „slika“ model i generirane slike sprema za daljnje korištenje u treniranju. Kamera se tijekom kretanja ne rotira po prednjoj osi nego se rotira oko objekta po zadanim koracima.

Za rad su korišteni modeli koji su ilustrirani u Prilogu A. Cilj je bio koristiti modele koji imaju raznolike stupnjeve sličnosti, više o tome u potpoglavlju 3.2.1.

### 3.2. Usporedba postojećih skupova podataka

Na internetu uvijek je izazov pronaći kvalitetan skup podataka za svrhu projekta ili rada. Jedan od izazova spomenutih u potpoglavlju 1.1. je u vezi nedostatka podataka za treniranje, ono se veže dalje ovdje. Naime, ako postoji nedostatak potrebno je: dobiti više podataka iz potrebnog izvora (npr. više označenih slika iz područja medicine, kao što su slike magnetske rezonancije MRI, podatci dobiveni iz EKG-a od pacijenata sa specifičnim bolestima, rendgen slike kostiju za svrhe otkrivanja specifičnih bolesti, itd.) ili je potrebno te podatke „krivotvoriti“. Npr. potrebno je iskoristiti proširenje skupa podataka ili sintetičko generiranje.

Umjetni podatci imaju široki raspon svrha za koje se mogu koristiti [21] i najčešće pojednostavljuju te pojeftinjuju operacije za koje su vezane. Za slučaju ovog rada se ne mora par dana slikati LEGO kocke, nego je moguće generirati sve slike u raznim željenim uvjetima. Mogu biti s raznim pozadinama, mogu dijelovi kocaka biti skriveni, svijetlo na kockama se može namještati, slike se mogu generirati oštećene kao simulacija slikanja u pravom svijetu (primjer je generirana slika s pomućenjem u nekom smjeru, kao da je osoba koja je slikala micala kameru pri slikanju, engl. *motion blur*).

Jedan od sintetičkih skupova podataka na internetu je od prethodno spomenutog Joost Hazelzet sa stranice Kaggle [19]. Podatkovni skup sadrži 50 klasa što je jednako 50 modela LEGO kocaka, gdje je svaki model uslika 800 puta iz raznih kutova. Jedan od problema sa Hazelzetovim skupom jest da su sve slike neprozirne, sadrže crnu pozadinu. Neprozirnim slikama je teže dodati prozirnost u svrhe generalizacije sa pozadinama i zbog toga je odlučeno kreirati novi skup podataka. Nad Hazelzetovim skupom podataka su neki Kaggle korisnici napravili mreže za klasifikaciju, kao što je DenseNet201 mreža korisnika stpete\_ishii na [22] s validacijskom točnošću od 66,35% i MobileNet mreža korisnika Datalira na [23] s validacijskom točnošću od 85,69%.

Još jedan skup podataka jest od prethodno spomenutog Daniel Westa, West tvrdi da je kreirao skup podataka od preko 100 000 slika prema [24], iako taj skup nije javno objavljen.

#### 3.2.1 Generirani skup podataka i ostvareni rezultati

Sintetički generirani podatci, za slučaj u radu sa LEGO kockama, imaju oštro definirane oblike. Kocke imaju rubove i dugmadi, ploče su tanjeg profila, štapići za povezivanje su

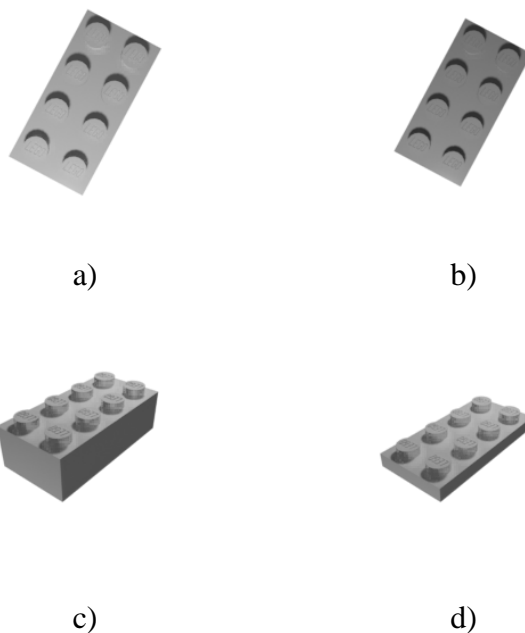
cilindričnog oblika sa specifičnim reljefima na krajevima, kotači su široki cilindri s reljefima na kontaktnoj površini, itd.

Na gore spomenutoj github stranici rada se nalaze *Jupyter* bilježnice s modelima, njihove strukture i rezultati preko ispisa u bilježnicama.

**Tablica 3.1.** Prikaz rezultata za korištene modele

| Model        | loss   | acc    | val_loss | val_acc |
|--------------|--------|--------|----------|---------|
| test         | 2,5988 | 15,80% | 2,4492   | 17,83%  |
| test_complex | 1,0020 | 66,06% | 0,9606   | 63,97%  |
| test_vgg16   | 0,3314 | 88,30% | 0,3880   | 85,97%  |
| test_vgg19   | 0,3692 | 87,06% | 0,4012   | 84,38%  |

Nad prethodno generiranim sintetičkim podacima jednostavne mreže ne mogu postići visoku točnost kao što je vidljivo na tablici 3.1., jednostavnije mreže nemaju dovoljno slojeva konvolucije da mogu otkriti potrebne značajke. Problem na koje nailaze kompliciranije mreže je vezan za sličnosti među modelima. Uzmimo za primjer model 3001 2x4 kocku i model 3020 2x4 ploču (slike se nalaze u prilogu A, konkretan primjer na slici 3.3.). Ako se mreža trenira sa slikama modela od skroz gore ili dolje, mreža će naučiti da su to značajke ta dva modela i ažurirat će težine prema tome. Ali kad dođe u pitanje predviđanje modela takve slike, mreža neće moći s potpunom sigurnošću reći koji je to model osim ako nije slikano na takav način da je dovoljno vidljivo sa strane, da se vide druge strane modela. Sličan koncept se može primijeniti i na model 3037 2x4 kocke za krov, ako se da slika od dolje to je još jedan sličan model za mrežu i tu postoji vjerojatnost da mreža napravi grešku u predviđanju, zbog toga nije moguće da mreža postigne neku vrlo visoku sveukupnu točnost. Paralelno tome je odabrano još par modela koji su ili slični ili jedinstveni u izgledu, npr. model 2340 kormilo i 4083 ograda, ti modeli su iz skoro svih pogleda relativno jedinstveni, dok postoje dodatni parovi za otežavanje predviđanja mreže kao što su model 2357 1x2x2 rub kocka i model 2420 1x2x2 rub ploča, gdje je jedina razlika u debljini modela.



**Slika 3.3.** Prikaz modela 3001 i 3020 za usporedbu poteškoće predviđanja klase modela, a) i b) prikazuju pogled odozgo dok c) i d) prikazuju pogled sa strane gdje se vidi razlika

Jedno moguće rješenje ovog problema je uvođenje „druge kamere“, tj. dodati još jedan unos u neuronsku mrežu [25]. Ako mreža ima pristup dvjema slikama istog modela, može postići bolju točnost prilikom predviđanja modela u slikama. Mrežu će na ovaj način biti teže koristiti u pravom životu ako je dostupna samo jedna kamera (npr. od mobilnog uređaja), ali lakše ako postoji stroj za specifično ovu svrhu ili ako je izrađena aplikacija koja slika model više puta, pa putem glasovanja odlučuje klasu.

Ako se uzme u obzir nesigurnost predviđanja mreže iz prošlog paragrafa, očekuje se predviđena maksimalna točnost od oko 85 do 90% za 25 klasa dok najbolje aktualno trenirane mreže [24] su postigle točnost od oko 90% (relativna točnost na više od 25 klasa), što podupire raspon postotka.



### 3.3. Implementacija ANN za klasifikaciju

U radu se koriste mreže VGG16 i VGG19 [26] za svrhe prijenosnog učenja. Aktivacijske funkcije umjetnih neurona spomenutih u potpoglavlju 1.1. su sve *ReLU* funkcije koje su često korištene u praksi. *ReLU* (engl. *Rectified Linear Unit*, ispravljena linearna jedinica) i njena derivacija sadrže sljedeći oblik:

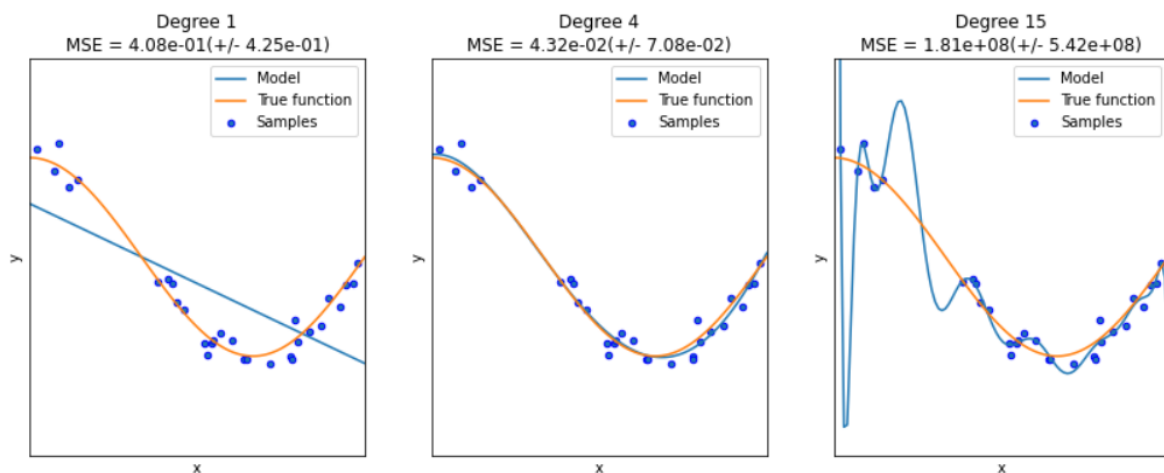
$$f(x) = \max(0, x), \quad f'(x) = \begin{cases} 0 & \text{ako } x < 0 \\ 1 & \text{ako } x > 0 \\ \text{neodređeno} & \text{ako } x = 0 \end{cases} \quad (3-1)$$

gdje  $x$  je ulazna vrijednost, a  $f(x)$  izlazna vrijednost funkcije.

Aktivacijska funkcija mora biti derivabilna radi svrhe treniranja algoritmom povratnog rasprostiranja pogreške (engl. *backpropagation algorithm*) [27]. Na kraju je potrebna optimizacijska funkcija, odabran je Adam algoritam koji je oblik algoritma stohastičkog gradijentnog spusta iliti SGD [28]. Adam algoritam je nadogradnja *RMSProp* [29] algoritma na takav način da ukomponira svojstva *AdaGrad* [30] algoritma. Umjesto da samo bazira stopu učenja za svaki parametar prema srednjoj vrijednosti parametara, ono uzima u obzir i adaptivni gradijent *AdaGrad* algoritma u obliku momenta gradijenta. Taj moment pomaže algoritmu ubrzati pronalazak optimalne težine za svaki parametar (ubrzava pronalazak točke konvergencije gradijentnog spusta). Adam zajedno sa ReLu aktivacijskom funkcijom se koristi pri treniranju dubokih neuronskih mreža s više skrivenih slojeva jer sadrži manju vjerojatnost kreiranja mrtvih neurona zbog prethodno spomenutih učinaka (mrtvi neuroni su onda samo oni čiji su parametri zapeli u nuli, jer nije moguć povratak iz nje).

Ciklus treniranja nad skupom podataka se naziva epoha, svaka epoha se provodi nad ulaznim skupom podataka iz kojeg se odabire mini-šarža (engl. *mini-batch*) podataka, koristeći Adam i povratno razmnožavanje ovaj se postupak iterativno ponavlja dok se ne iskoriste sve mini-šarže skupa podataka. Na kraju epohe se na validacijskom skupu podataka provjeravaju performanse mreže, ti parametri se onda mogu koristiti u sljedećim epohama za namještanje parametara treninga. Kad završi trenutna epoha pokreće se sljedeća dok ne prođu sve zadane epohe. Algoritam treniranja može rano zaustaviti treniranje ako vidi da se funkcija greške ne mijenja znatno preko vremena, ovo je naznaka da je algoritam treniranja došao u lokalni minimum za moguću grešku i da je potrebno dalje namještati parametre ili ako su zadovoljavajući rezultati spremi model za daljnje korištenje.

Daljnji problemi pri treniranju se mogu pojaviti u obliku pretreniranja (engl. *overfitting*). To se događa ako je skup podataka za treniranje vrlo malen, ili ako su podatci vrlo slični jedni drugima. Prilikom treniranja mreže potrebno je obratiti pažnju na parametre koje se vraćaju na kraju validacije, koje govore kako model u toj epohi djeluje nad validacijskim skupom. Ako model ima dobre performanse nad trening skupom ali loše performanse nad validacijskim skupom može se reći da pati od pretreniranja nad trening skupom. Potrebno je namjestiti parametre mreže ili namjestiti dani skup podataka na takav način da se smanji pretreniranje (više uzoraka, mijenjanje parametara proširenja skupa podataka).



**Slika 3.4.** Grafički prikaz za utjecaj parametara treninga na pretreniranje [31]

Na slici 3.4. se vide primjeri tri modela s drugačijim parametrima treniranja. Prvi model je oblika polinomna regresija s jednim stupnjem, zbog toga se ne može točno podacima namjestiti (engl. *underfit*). Drugi model ima četiri stupnja, što je dostatno da se model namjesti trening podacima i da liči originalnoj funkciji. Treći model ima previše stupnjeva i pokušava se namjestiti na takav način da kroz sve podatke pokušava proći. Tu se nalazi problem pretreniranja.

Sličan koncept se može dogoditi i kod neuronskih mreža, ako preko Adam i povratnog razmnožavanja konstantno dolaze slični gradijenti, neuroni će biti naučeni na te specifične podatke preko ažuriranja težina i kad dođu neki drugi dovoljno drugačiji podatci, neuroni neće biti u sposobnosti dati kvalitetan odgovor prema sljedećem sloju, što uzrokuje grešku u npr. predviđanjima i pogoršava parametre tijekom validacije. Otkrivanje podtreniranja i pretreniranja je jednostavno ako imamo parametre treniranja, ako model ima dobre performanse prilikom treniranja i dobre performanse pri validaciji, onda se može reći da model kvalitetno napreduje u treniranju. Ako model ima loše performanse kod validacije onda je

moгуće pretreniran, dok loše performanse kod treniranja ukazuju da nije dovoljno treniran, tj. podtreniran je i nedostaje više epoha treniranja.

### 3.3.1. Opis slojeva i strukture

Potrebno je učitati podatke zajedno sa svim parametrima vezano za proširenje skupa podataka, kao što je prikazano na slici 3.5.

#### *Linija    Kod*

```
1:      import numpy as np #prvo potrebni uvozi
2:      import keras
3:      import matplotlib.pyplot as plt
4:      import cv2
5:      from keras.applications.vgg16 import VGG16
6:      from keras.preprocessing.image import ImageDataGenerator
7:      from keras.models import Model
8:      dataset_path = "../..//LEGO_brick_custom/"
9:      dataset_output = "./output"
10:     bg = Background_source("../..//background_images", (200,200),
11:     (3,3))#kreiranje objekta za postavljanje pozadinskih slika
12:     train_datagen = ImageDataGenerator(
13:         shear_range=10,
14:         zoom_range=[1,1.2],
15:         height_shift_range=0.1,
16:         width_shift_range=0.1,
17:         brightness_range=[0.6,1],
18:         rotation_range=90,
19:         horizontal_flip=True,
20:         vertical_flip=True,
21:         preprocessing_function = bg.apply_background_preprocess_input
22:     )
23:     validation_datagen = ImageDataGenerator(
24:         preprocessing_function = bg.apply_background_preprocess_input
25:     )
26:     train_generator =
27:     train_datagen.flow_from_directory(os.path.join(dataset_output,'train')
28:     color_mode="rgba", target_size=(200,200), shuffle=True,
29:     class_mode="categorical")
30:     validation_generator =
31:     validation_datagen.flow_from_directory(os.path.join(dataset_output,
32:     'val'), color_mode="rgba", target_size=(200,200), shuffle=True,
33:     class_mode='categorical')
```

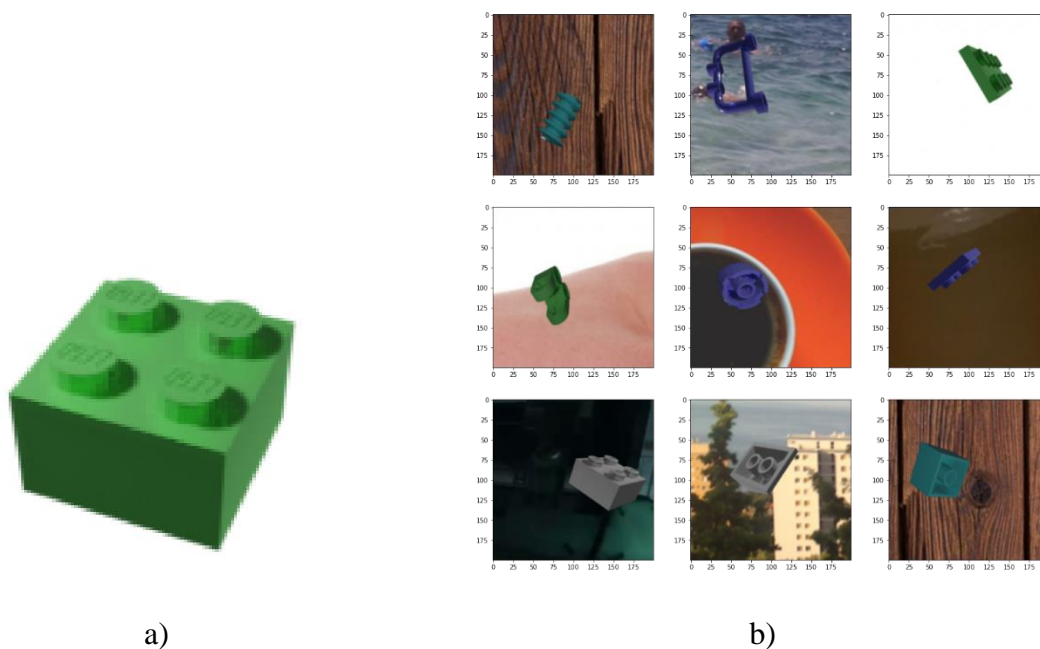
**Slika 3.5.** Blok koda za uvoz podataka i označavanje vrste proširenja skupa podataka

U kodu su odabrani sljedeći parametri u vezi proširenja skupa podataka:

- Nasumično smicanje: raspon od  $-10^{\circ}$  do  $10^{\circ}$
- Nasumično povećanje (engl. *zoom*) slike, raspon od 0% do 20%
- Nasumična translacija horizontalno, raspon od -10% do 10%

- Nasumična translacija vertikalno, raspon od -10% do +10%
- Nasumično mijenjanje svjetline, raspon od 50% do 100% originalne svjetlosti
- Nasumična refleksija horizontalno i vertikalno
- Nasumična rotacija, raspon od  $0^\circ$  do  $90^\circ$

Koristeći refleksiju u oba smjera zajedno s nasumičnom rotacijom od  $90^\circ$ , pokrivene su sve mogućnosti pozicije nekog LEGO modela, zbog ovog razloga kamera nije rotirana tijekom generiranja sintetičkih podataka. Dobivaju se rezultati s prihvatljivim izgledom nad osnovnim slikama s prozirnošću kao na slici 3.6. a). Koristeći prethodne postavke za proširenje skupa podataka dobivaju se slike kao po slici 3.6. b) s nasumičnom pozadinom za bolju generalizaciju i smanjivanje razlike u domenama.



**Slika 3.6.** a) Slika modela 3003 LEGO kocke, b) Slika modela LEGO kocaka proširene raznim tehnikama

Odabrana najbolja mreža po tablici 3.1. je duboka umjetna neuronska mreža zasnovana na prijenosnom učenju VGG16 mreže. Mreža je kreirana po slici 3.7.

#### *Linija    Kod*

```
1:     from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense,
      Dropout, Lambda, Input
2:     def remove_alpha(image):
3:         return image[:,:,:,:-1]
4:
5:     new_input = Input(shape=(200,200,4))
6:     lam = Lambda(remove_alpha, output_shape=(200,200,3))(new_input)
7:     temp = Model(inputs=new_input, outputs=lam)
8:
9:     base_model = VGG16(weights='imagenet', include_top=False,
      input_tensor=temp.output)
10:    for layer in base_model.layers:
11:        layer.trainable = False
```

**Slika 3.7.** Blok koda za uvoz i kreiranje VGG16 arhitekture

Zato što je arhitektura VGG16 zasnovana na neprozirnim slikama, a u radu se rukuje sa slikama koje sadrže četvrti kanal, alfa kanal za prozirnost. Potrebno je na samom ulazu maknuti taj alfa kanal (slika 3.7. linija 2, linija 6) koji se dobiva iz generatora (slika 3.5.) da bi sve pravilno radilo s postojećim težinama VGG16 mreže jer alfa kanal je napunjen samo jedinicama, nema vrijedne značajke u sebi.

Prijenosno učenje se izvodi tako da se parametar *include\_top* u osmoj liniji slike 3.7. postavi na *false* dok se osnovne težine preuzmu s interneta koristeći *weights='imagenet'*. Nakon toga je potrebno postaviti novi blok gusto spojenih neurona (slika 3.8.) da služi kao novi klasifikator mreže, ono koristi izlaze postojećih konvolucijskih blokova da bi odlučilo o kakvoj kocki na slici se radi.

#### *Linija    Kod*

```
1:     top_model = base_model.output
2:     top_model = Flatten()(top_model)
3:     top_model = Dense(1024)(top_model)
4:     top_model = Dropout(0.5)(top_model)
5:     top_model = Dense(512)(top_model)
6:     top_model = Dense(len(train_generator.class_indices),
      activation="softmax")(top_model)
7:
8:     model = Model(inputs = base_model.inputs, outputs = top_model)
```

**Slika 3.8.** Blok koda za kreiranje novog bloka neurona za klasifikaciju

U slici 3.8. se vidi kod za postavljanje odgovarajućih slojeva koji čine novi blok gusto spojenih umjetnih neurona. Sloj ispada (engl. *dropout*) se koristi za regularizaciju, jednostavno rečeno ono služi kao most između dva sloja i ovisno o zadanom postotku, toliko posto konekcija blokira tijekom jedne epohe treniranja. Po [32] smanjuje se pretreniranje i ne utječe toliko jako na stabilnost mreže.

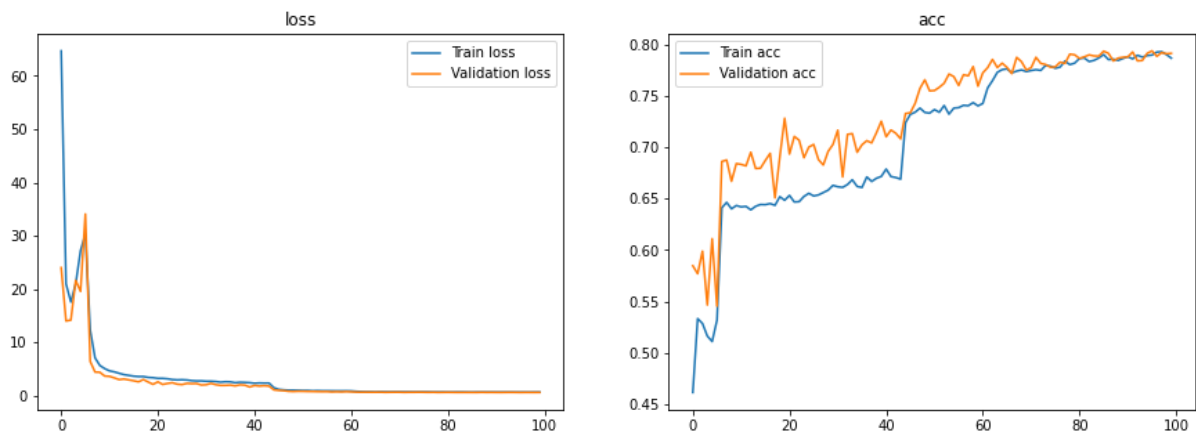
Na kraju je potrebno izvesti treniranje u dva koraka, prvi put sa smrznutim slojevima originalne mreže (provedeno u slici 3.7.). Da su ti slojevi otključani, nasumične težine u novom gusto spojenom bloku bi uzrokovale razna, i po mogućnosti štetna, ažuriranja težina u konvolucijskim blokovima, što može slijediti do uništenja postojećih detektora rubova i oblika koje je mreža naslijedila od originalne VGG16 mreže.

#### *Linija    Kod*

```
1:     from keras.optimizers import Adam
2:     opt = Adam(lr=1e-3)
3:     model.compile(
        loss='categorical_crossentropy',
        optimizer=opt,
        metrics=['accuracy']
    )
4:
5:     history = model.fit_generator(
        train_generator,
        validation_data = validation_generator,
        epochs = 100,
        verbose = 1,
        callbacks=callbacks
    )
```

**Slika 3.9.** Blok koda za prvo treniranje

Tijekom treniranja u Jupyter bilježnici se postavlja ispis. Time se može pratiti trenutačno stanje treniranja i na kraju epohe prikazuje vrijednost funkcije gubitka i točnosti tijekom treniranja i validacije. Po tim metrikama se mjeri kvaliteta mreže. Nakon prvog treniranja se može pregledati *history* varijabla preko grafičkog ispisa.



**Slika 3.10.** Grafički prikaz metrika preko epoha tijekom prvog treniranja

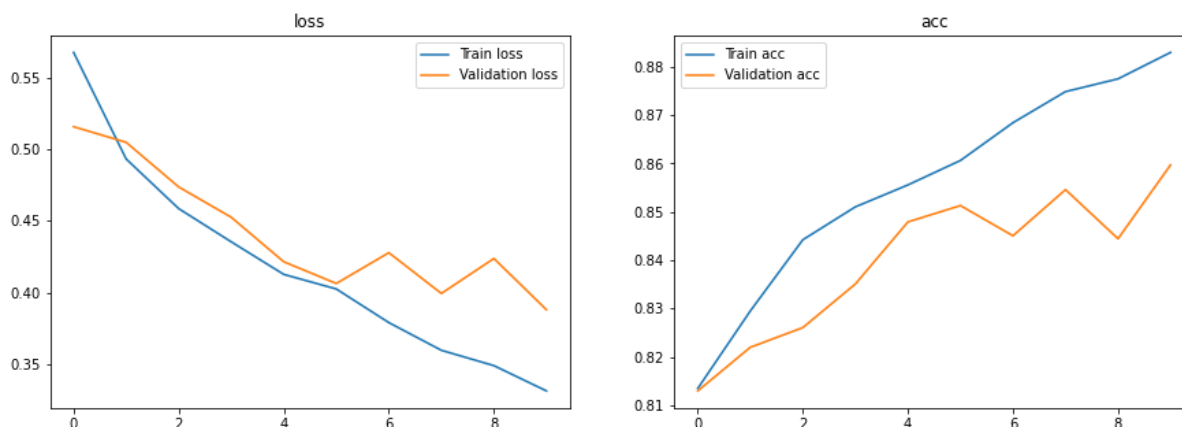
Po grafu na slici 3.10. se vidi da su pravilno odabrani parametri za stopu učenja u Adam optimizacijskom algoritmu, iako u treniranju sudjeluje i adaptivno mijenjanje stope učenja preko povratnih poziva (engl. *callback*) na takav način da ako se ne dobiju bolji rezultati metrika unutar 4 epoha, smanji stopu učenja na 25% trenutne vrijednosti i nastavi treniranje. Adaptivno smanjenje stope učenja dovodi do bolje konvergencije Adam algoritma. Drugo učenje se provodi tako da se omogući treniranje nad zadnjim konvolucijskim blokom.

#### Linija Kod

```
1:     for layer in base_model.layers[:-4]:
2:         layer.trainable = True
3:     opt = Adam(lr=1e-6)
4:     model.compile(
5:         loss='categorical_crossentropy',
6:         optimizer=opt,
7:         metrics=['accuracy']
8:     )
9:     history = model.fit_generator(
10:         train_generator,
11:         validation_data = validation_generator,
12:         epochs = 10,
13:         verbose = 1,
14:         callbacks=callbacks
15:     )
```

**Slika 3.11.** Blok koda za drugo treniranje

Potrebno je odmrznuti zadnja četiri sloja (zadnji konvolucijski blok) kao na slici 3.11. i postaviti optimizacijski algoritam na nisku početnu vrijednost. Dobiveni rezultati su grafički prikazani na slici 3.12.



**Slika 3.12.** Grafički prikaz metrika preko epoha tijekom drugog treniranja

Vidljivo je poboljšanje jer ovaj put je blok gusto povezanih neurona naviknut na izlaz konvolucijskih mreža i neće odmah uništiti konvolucijski blok, nego se daje prilika konvolucijskom bloku da se privikne na izlaze prethodnih konvolucijskih blokova u slučaju učenja nad LEGO kockama.

Matrice zabune (engl. *confusion matrix*) su grafički prikaz svih predviđanja koje je neka mreža izvela [33]. U ovom slučaju gdje imamo umjetnu neuronsku mrežu s više klasa, dobiti će se matrica zabune  $N \times N$  polja, gdje  $N$  stoji za broj klasa. U vertikalnoj osi svaki redak predstavlja istinitu oznaku (engl. *true label*), dok na horizontalnoj osi svaki stupac predstavlja klasu s njenim predviđenim brojem pogodaka (engl. *predicted label*). „Cilj“ mreže bi bio što više polja u dijagonali „pogoditi“ jer ta polja su istiniti pozitivni (engl. *true positive*), dok svi ostali su lažni pozitivni (engl. *false positive*).

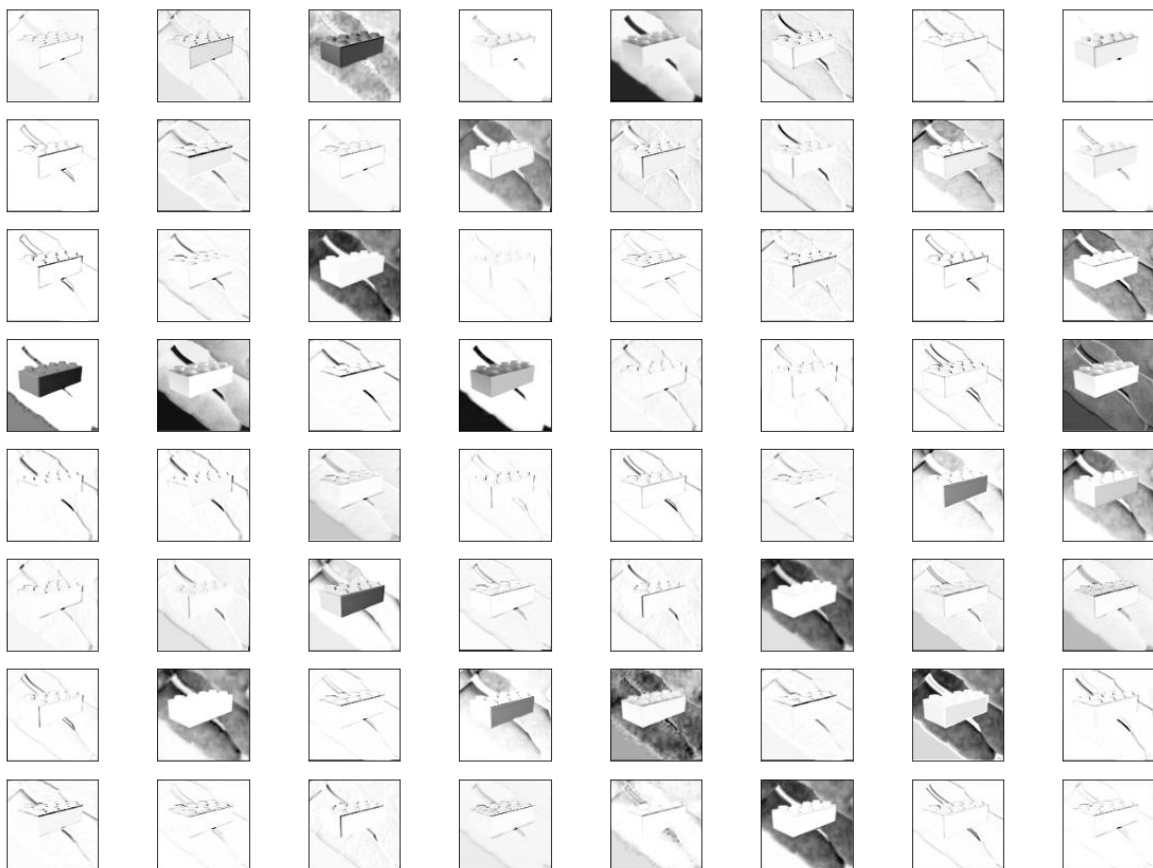
Iz matrice zabune najboljeg modelu ovog rada (slika 3.13.) se može jasno vidjeti greška zabune gdje umjetna neuronska mreža pogrešno odabere klasu jer je previše slična iz nekog kuta nekoj drugoj klasi.





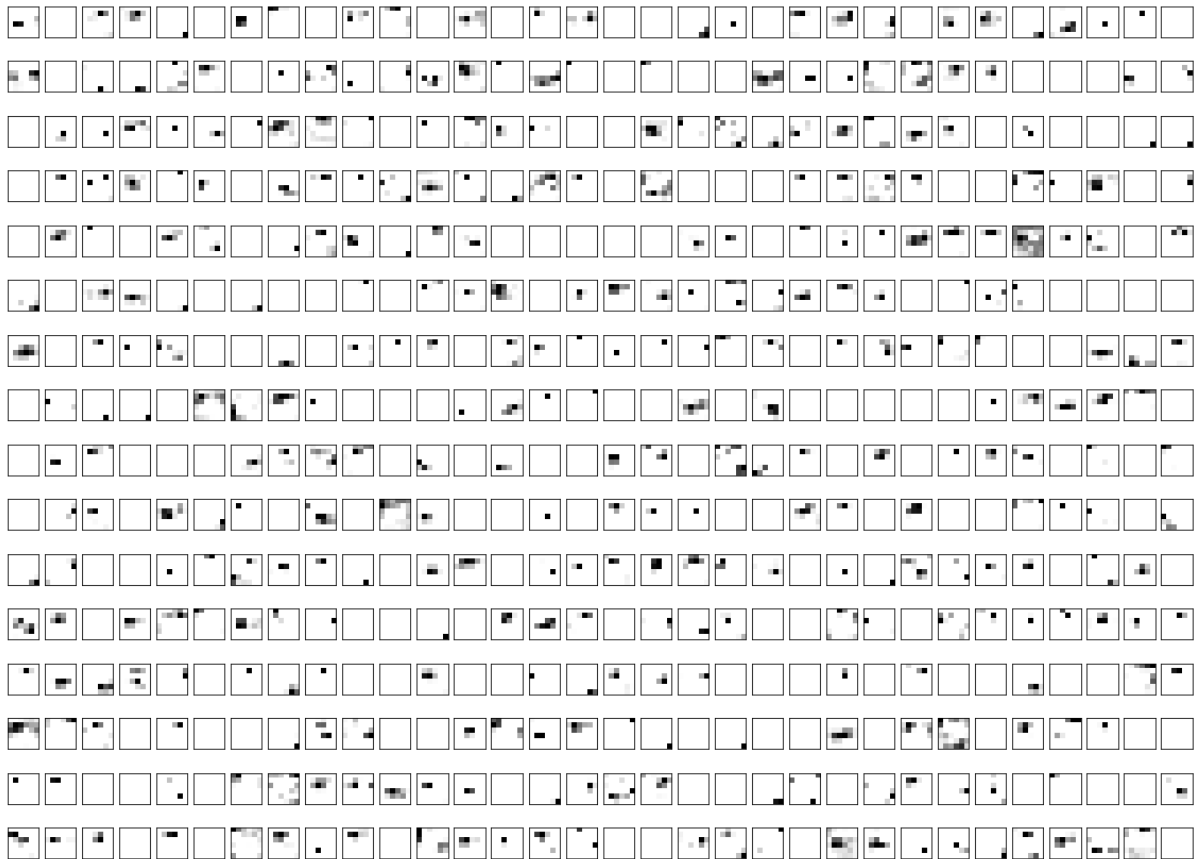
relativno visok broj podudaranja s drugim sličnim klasama kao što je 2x4 kocka (101 lažnih pozitiva da je 2x4 kocka) i LEGO kocka 2x4 sa 45° nagibom (14 lažnih pozitiva). Očekivalo bi se da je broj pogrešnih podudaranja sličan za te dvije klase, iako postoji tek toliko slika odozgo. Ovime se vidi iz matrice da je umjetna neuronska mreža osjetljiva tek na sličnosti među klasama (tj. modelima LEGO kocaka) i da nasumična pozadina ne utječe puno na njenu sposobnost predviđanja klase. Jedan mogući problem matrice zabune je u tome što ono prikazuje predviđanje s najvećim postotkom vjerojatnosti, ne radi sa svim decimalnim težinama za neki bolji prikaz sličnosti između klasa. Ono tek toliko podataka može staviti u prikladan prostor da bude prihvatljiva vidljivost danih podataka.

Dodatan alat za analizu kvalitete umjetne neuronske mreže je prikaz slika značajki (engl. *feature maps*). Slike značajki se uzimaju se iz izlaza konvolucijskih slojeva da bi dobili prikaz toga na što im se izlazi svih filtara aktiviraju, kao što je prikazano na slici 3.14.



**Slika 3.14.** Prikaz slika značajki na izlazu prvog konvolucijskog bloka

U promatranoj umjetnoj neuronskoj mreži se samo zadnji konvolucijski blok trenira, i to na kraju treninga joj se otključa sposobnost treniranja na 10 epoha. Međutim, izlaz tog konvolucijskog bloka je 512 slika značajki s rezolucijom 6x6 piksela, iz čega se ne može vidjeti značajna informacija ali je dobro vidjeti ikakvu aktivaciju nad slikama filtera.



**Slika 3.15.** Prikaz slika značajki na izlazu zadnjeg konvolucijskog sloja

Koristeći informaciju dobivenu iz svih konvolucijskih blokova koristeći njihove slike značajki, moguće je zaključiti da umjetna neuronska mreža sadrži dobru sposobnost otkrivanja potrebnih rubova i oblika koje je naslijedila od VGG16 mreže. Aktivnost u zadnjem konvolucijskom bloku dokazuje da konvolucije prethodnih blokova izvlače potrebne značajke da se tom zadnjem bloku aktiviraju izlazi i time daju potrebni podatci za predviđanje u bloku potpuno povezanih neurona.

### 3.4. Usporedba s drugim umjetnim neuronskim mrežama

Dobivena umjetna neuronska mreža za prepoznavanje LEGO kocaka se može koristiti samo s originalnim Python okruženjem u kojem je napravljen zbog specifičnih verzija *Keras* i *Tensorflow* biblioteka koje su korištene za *PlaidML* sloj abstrakcije. Zbog toga postoje poteškoće sa prijenosom mreže među sustavima jer osim spremljene arhitekture i težina slojeva

je potreban popis biblioteka koje *Python* koristi sa točnim verzijama, što je potrebno držati u umu ako je poželjno testirati mrežu danu u *github* poveznici.

Treniranje dobivene mreže je relativno sporo zbog bloka gusto povezanih neurona koje mora 25 klasa predvidjeti iako neke druge mreže mogu imati iznad 1000 klasa, što zahtjeva više neurona za kvalitetno predviđanje. Npr. originalni VGG16 je sadržavao dva sloja od 4096 neurona i konačan sloj od 1000 neurona dok dobivena mreža koristi dva sloja od 1024 i 512 neurona te konačan sloj od 25 neurona. Za treniranje je bilo potrebno 22 sata za 100 epoha prvog treniranja sa smrznutim konvolucijskim blokovima dok je za drugo treniranje bilo potrebno oko 5.5 sati za 10 epoha drugog treniranja s odmrznutim zadnjim konvolucijskim blokom. Sveukupno oko 27.5 sati treniranja nad 38 400 slika LEGO kocaka u 25 klasa, što iznosi ~1400 obrađenih ulaza svaki sat iliti ~0.38 ulaza svake sekunde. Osim toga što sadrži visok broj neurona u mreži, procesor može biti zauzet dok generira proširenje podataka nad ulazom (afine transformacije pa nakon toga dodavanje pozadine). Zahvaljujući paralelizmu uvedenom u *Kerasu* između treniranja i generiranja podataka ovaj se utjecaj ne osjeti toliko jako u jačim strojevima koji to mogu brže procesirati.

Ankita Singh i Pawan Singh iz Indijskog sveučilišta Amity su napravili pregled polja prepoznavanja slika koristeći umjetne neuronske mreže [34]. U radu su opisali neuronske mreže i njihove koristi te korake izrade neuronske mreže. Nad CIFAR-10 [35] skupom podataka su kreirali umjetnu neuronsku mrežu (6000 slika za 3 klase: zrakoplovi, ptice i automobili). Umjetna neuronska mreža se zasnivala na tri bloka konvolucije i jedan blok gusto povezanih neurona. Blok gusto povezanih neurona je sadržavao jedan sloj od 3 neurona koji je primao 512 ulaza prethodnog konvolucijskog bloka. Treniranje se naspram dobivene mreže izvelo puno brže, sa brzinom od 4-5 milisekundi po ulazu, što je ~222 ulaza po sekundi (oko 580 puta brže).

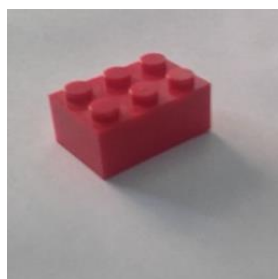
Razlika među mrežama se nalazi u tome što dobivena mreža sadrži masivan broj parametara (težina) koje se moraju ažurirati tijekom treniranja, dok u Singhovoj mreži broj parametara je manji (512 na 3 neurona naspram našeg 18432 na 1024 na 512 na 25 neurona, i to samo za gusto povezani blok gdje se nalazi velika većina parametara).

Sličnu priču možemo navesti sa velikim mrežama kao što je originalni VGG16, gdje postoji puno veći broj parametara za svrhe klasifikacije više klasa (18432 na 4096 na 4096 na 1000), gdje će vrijeme treniranja trajati duže nego kod dobivene mreže i puno duže od Singhove mreže.

## 4. DEMONSTRACIJA I ISPITIVANJE FUNKCIONALNOSTI SUSTAVA

### 4.1. Demonstriranje mogućnosti sustava

Dobivena duboka umjetna neuronska mreža može nad 200x200 slikama početi prepoznavati objekte ako se ti objekti nalaze unutar prethodno definiranih i treniranih 25 klasa. Ne postoji mogućnost pravilnog prepoznavanja ostalih modela LEGO kocaka. Npr. mreža ne zna prepoznati LEGO 3x2 kocku 3002 jer ju uopće nema kao mogućnost na izlazu i nema težine koje su ažurirane takvim ulazima. Ako se mreži da slika tog modela, dobiti će se nasumične vrijednosti predviđanja, često najbližeg modela (npr. dobije se predviđanje slike 4.1. po slici 4.2.).



**Slika 4.1.** LEGO kocka model 3002

*Linija    Kod*

```
1:     image_path = "../../../test_images/3002.png"
2:     image = cv2.imread(image_path, cv2.IMREAD_UNCHANGED)
3:     image = bg.apply_background_preprocess_input(image)
4:     image = np.expand_dims(image, axis=0)
5:     prediction = loaded_model.predict(image)
6:     np.set_printoptions(formatter = {'float': lambda x:
7:     print(prediction[0])
8:     print(prediction[0].tolist().index(max(prediction[0])))
9:     print(list(validation_generator.class_indices.keys())[prediction[0]
10:    .tolist().index(max(prediction[0]))])
...
ispis: [0.00 0.00 0.00 0.00 0.92 0.00 0.00 0.00 0.02 0.00 0.00 0.00 0.00
0.05 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01]
4
3001 Brick 2x4
```

**Slika 4.2.** Blok koda za pripremu predviđanje i ispis predviđanja

Takva predviđanja dobivena mreža ne može očekivati jer neuroni nisu „naučeni“ na izlaze konvolucijske mreže za LEGO kocku model 3002 („predvidjelo“ je LEGO kocku model 3001).

Ako se na sličan način preda slika 2357 modela, za koju je mreža trenirana, dobiju se drugačiji rezultati.



**Slika 4.3.** LEGO kocka model 2357

*Linija    Kod*

```
ispis: [0.00 0.00 0.36 0.04 0.00 0.32 0.01 0.00 0.01 0.08 0.06 0.00 0.00
0.00 0.00 0.00 0.08 0.00 0.00 0.00 0.00 0.02 0.00 0.00 0.01]
2
2357 Brick corner 1x2x2
```

**Slika 4.4.** Ispis za predviđanje slike 4.3.

Na slici 4.4. je prikazan ispis predviđanja svih klasa i kojoj klasi ono misli da slika pripada. Postoji puno mogućih smetnji u slici i zato je predvidjelo s niskim postotkom od 36% da je 1x2x2 rub kocka, a ne sljedeći najbliži 2x2 kocka. Postoji mogućnost da sustavu smeta sjena kocke jer trening podatci ne sadrže sjenu osim na dugmadima.

Za sliku 4.5. mreža kaže s puno većom pouzdanošću da se radi o LEGO kocki model 3001.



**Slika 4.5.** LEGO kocka model 3001

*Linija    Kod*

```
ispis: [0.00 0.00 0.00 0.00 0.96 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00
0.03 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01]
4
3001 Brick 2x4
```

**Slika 4.6.** Ispis za predviđanje slike 4.5.

Jedna mogućnost boljeg rada na ovoj slici je razlika objekta i pozadine, sa svijetlim objektom i tamnijom pozadinom, jer dobiva pouzdanost od 96% za klasu 3001. Ovo pravilo ne vrijedi uvijek, jer na sljedećem primjeru je mreža loše ocijenila o kojoj se kocki radi.



**Slika 4.7.** LEGO kocka model 3003

*Linija    Kod*

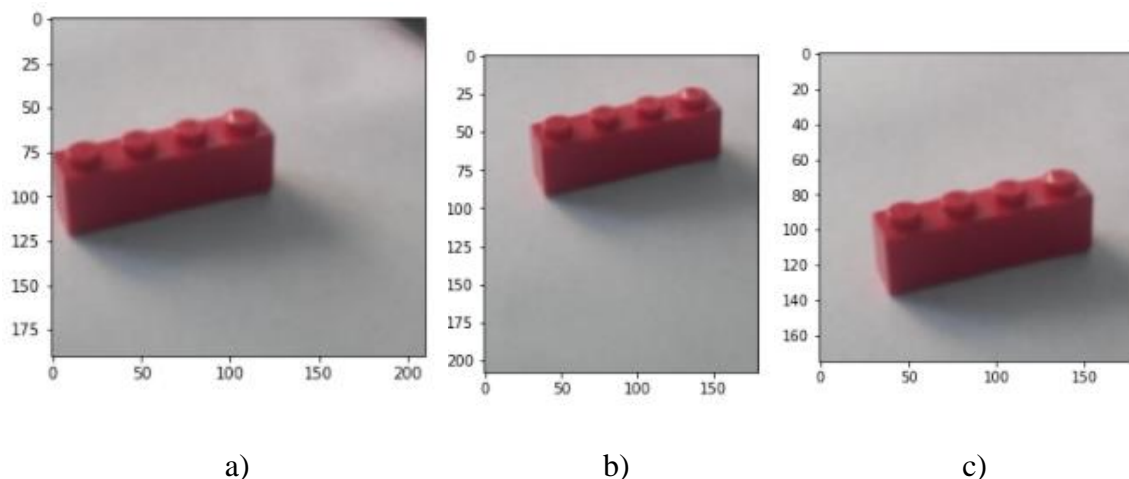
```
ispis: [0.00 0.00 0.00 0.00 0.27 0.12 0.00 0.00 0.00 0.00 0.00 0.00 0.00
0.18 0.00 0.00 0.40 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02]
16
3298 Roof tile 2x3
```

**Slika 4.8.** Ispis za predviđanje slike 4.7.

Točna klasa je dobila tek 27% vjerojatnosti, dok najviše ocijenjena klasa ima 40%. Razlika od 13% ali kad se gleda najveća vjerojatnost mreža je loše predvidjela.

## 4.2. Ispitivanje funkcionalnosti sustava

Dobivena mreža bi, po teoriji, trebala biti otporna na pomake objekta po slici i promijeni boje. Tijekom stvaranja ulaznih slika preko *ImageDataGenerator* u slici 3.5. dani su argumenti za afine transformacije, koje uključuje i pomak horizontalno i vertikalno. Otpornost na boje bi trebalo doći iz činjenice da se koriste nasumične boje u generiranju sintetičkih podataka, u nadi da će dobivena mreža naučiti ignorirati boje i koncentrirati se na rubove i oblike. U sljedeća tri primjera će se istražiti dali je dobivena mreža otporna na takve promjene.



**Slika 4.9.** LEGO kocka 3010 a) pomaknuta u lijevo, b) pomaknuta prema gore desno, c) pomaknuta prema dolje

*Linja*    *Kod*

```
ispis      [0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.98 0.00 0.00 0.00 0.00
a):        0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00]
8
3010 Brick 1x4
```

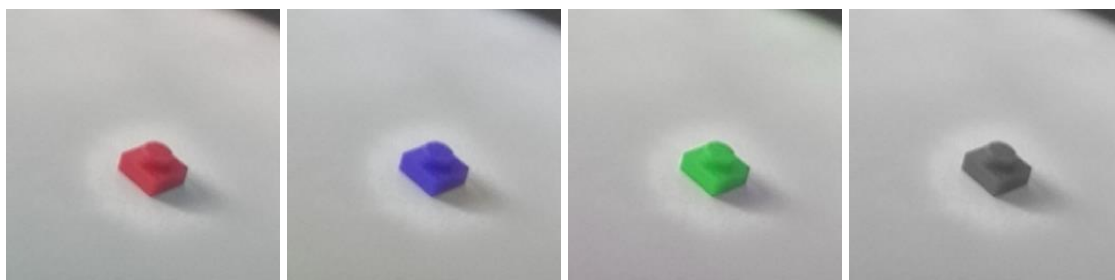
```
ispis      [0.00 0.00 0.00 0.00 0.04 0.00 0.00 0.01 0.87 0.00 0.00 0.00 0.03
b):        0.04 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00]
8
3010 Brick 1x4
```

```
ispis      [0.00 0.00 0.00 0.00 0.08 0.00 0.00 0.00 0.77 0.02 0.00 0.00 0.00
c):        0.13 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00]
8
3010 Brick 1x4
```

**Slika 4.10.** Ispis za predviđanje slika a), b) i c) u slici 4.9.

Dobivena mreža daje visok postotak vjerojatnosti za pomaknute slike, 98% za sliku 4.9. a), 87% za sliku 4.9. b) i 77% za sliku 4.9. c).





a)

b)

c)

d)

**Slika 4.11.** LEGO kocka 3024 a) originalna boja, b) zamjena crvenog i plavog kanala, c) zamjena crvenog i zelenog kanala, d) slika pretvorena u sive tonove

#### *Linija    Kod*

```
ispis [0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.10 0.00 0.00 0.00 0.02 0.76
a):   0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11 0.00]
12
3024 Plate 1x1
```

```
ispis [0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.33 0.00 0.00 0.00 0.01 0.42
b):   0.00 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.23 0.00]
12
3024 Plate 1x1
```

```
ispis [0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.61 0.00 0.00 0.00 0.01 0.18
c):   0.00 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.17 0.00]
7
3005 Brick 1x1
```

```
ispis [0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.07 0.00 0.00 0.00 0.05 0.81
d):   0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.07 0.00]
12
3024 Plate 1x1
```

**Slika 4.12.** Ispis za predviđanje slika a), b), c) i d) u slici 4.10.

Za originalnu sliku je dobiven točan rezultat s 76% vjerojatnosti. Mijenjanjem crvenog i plavog kanala dobije se 42% vjerojatnosti što je i dalje najveći postotak predviđanja. Mijenjanje crvenog i zelenog kanala uzrokuje netočan rezultat gdje točna klasa dobiva samo 18% vjerojatnosti, ali predviđena klasa se razlikuje samo u debljini (predviđena je 1x1 kocka, a ne ploča). Slika u sivim tonovima je dobila najbolji rezultat sa 81% vjerojatnosti predviđanja.

### 4.3. Analiza rezultata

Sve slike korištene u 4. poglavlju su ručno uslikane. Uslikane su s 1 megapiksela rezolucijom za najgori mogući rezultat (moderne kamere, čak i u mobilnim uređajima podržavaju više od 10 megapiksela inače). Dobivaju se slike relativno loše kvalitete s lagano

zamućenim rubovima LEGO kocaka, što predstavlja izazov dobivenoj mreži da otkrije odgovarajuće rubove i oblike za predviđanje. Sve slike se nalaze u github repozitoriju.

U potpoglavlju 4.1. predstavljene su slike i njihova predviđanja preko dobivene mreže. Vidljivo je preko rezultata da se dobivene vjerojatnosti ne podudaraju s vjerojatnostima nad sintetičkim podacima. Slika 4.5. je postigla zadovoljavajući rezultat, dok slika 4.3. je postigla točan rezultat samo zato što je njena vjerojatnost najveća s 36% dok sljedeća klasa (2x2 kocka) je dobila 32% vjerojatnosti. Kod prevrnutе LEGO kocke 2x2 dobilo je pogrešan rezultat s time da točna klasa nije ni na drugom mjestu, prvo mjesto je netočna klasa 3298 s 40%, sljedeća klasa je netočna klasa 3001 s 27% vjerojatnosti, sljedeća je opet netočna klasa 3037 s 18%, i tek onda je točna klasa sa samo 12% vjerojatnosti. Moguće da je tu smetala boja LEGO kocke, u podacima za treniranje ne postoje narančaste kocke.

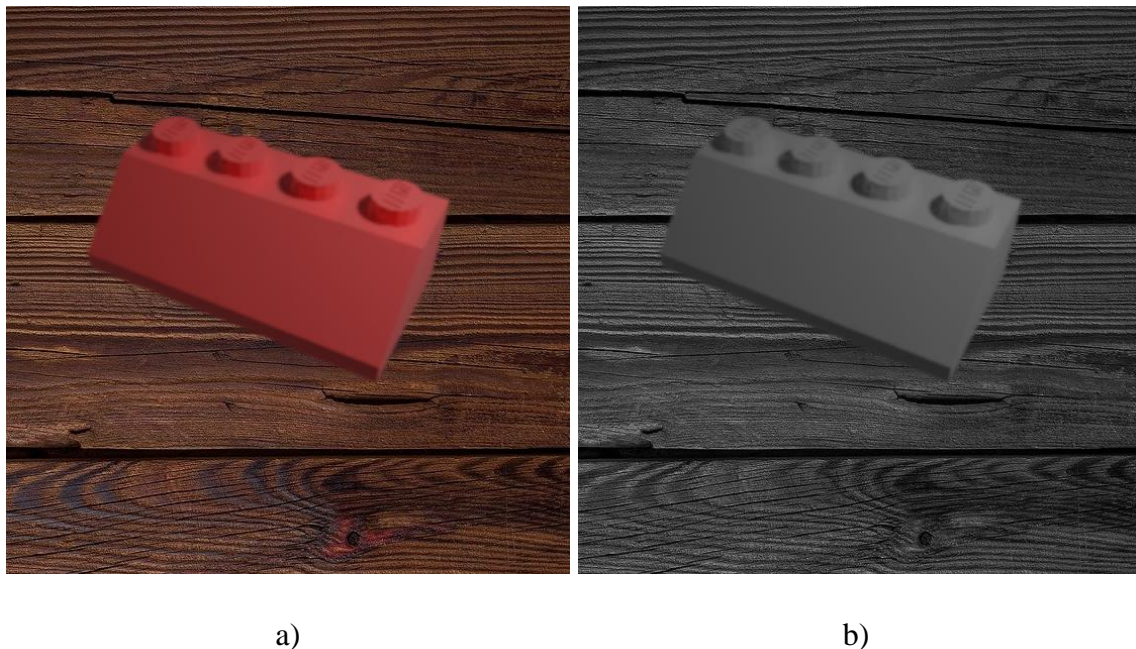
Dalje, u potpoglavlju 4.2. su korištene dvije jednostavne slike jasno vidljivih modela za provjeru otpornosti mreže nad pomicanjem gledanog objekta u okviru i promjena boja.

U prvom ispitivanju je uzeta slika LEGO kocke modela 3010 i pomicana je u tri smjera. U prvoj slici je diralo lijevi rub ali je i dalje dobro predvidjelo klasu LEGO kocke. U drugoj slici je pomaknuto prema gore, s time da se jedan dio slike odsiječe u procesu. To nije upitno za rad mreže jer joj ne smetaju nedostajući podatci na nekim od ulaza (podatci veći od ulaza ruše program). Druga slika je dobila lošije predviđanje, 87% naspram 98% vjerojatnosti prve slike. Treća slika je pomaknuta prema dolje, isto s odsječenim dijelom i ona je dala još lošije rezultate sa 77% dobivene vjerojatnosti.

Postoji mogućnost da nedostatak ulaza negativno utječe na predviđanje, ali to nema puno smisla jer prvo treba slika proći kroz 5 blokova konvolucije prije nego li može doći do blokova gusto povezanih neurona koje sadrže trenirane težine. Kod neurona bi smetalo nedostatak ulaza dok konvolucija i dalje radi bez problema jer kako se smanjuje veličina filtra a povećava broj filtara, smanjuje se utjecaj „izgubljenih“ podataka u kasnijim slojevima koji imaju filtre veličine 6x6, gdje je teško primijetiti takav nedostatak podataka.

Drugo ispitivanje je u vezi mijenjanja boja iste slike tijekom predviđanja. Dana je originalna slika LEGO kocke 3024 crvene boje za najveći učinak prilikom mijenjanja kanala boje. Originalna slika dobiva drugi najbolji rezultat od 76% vjerojatnosti. Sljedeće dvije slike su dobile 42% (crvena u plavu) i 18% (crvena u zelenu). Mijenjanje boje cijele okoline, a ne samo objekta, možda utječe na dobivenu mrežu. Lagano promijenjena boja papira kojoj su sada isto zamijenjeni kanali boja možda uzrokuje probleme tijekom predviđanja. Takve sitne razlike

nisu uopće vidljive ljudskom oku, ali umjetna neuronska mreža, koja rukuje s takvim podacima, može primijetiti te sitne razlike i dobiti utjecaj na predviđanje. Zadnja slika gdje je originalna slika pretvorena u sive tonove je dobila najbolje rezultate od svih danih slika za to ispitivanje s 81% vjerojatnosti. Moguće je da se rezultati mreže mogu poboljšati tako da se sve slike za treniranje pretvore u slike sivih tonova, ali tu postoji mogući problem detekcije rubova po boji. Npr. lakše je otkriti crvenu LEGO kocku na smeđem drvenom stolu kao po slici 4.13. a) nego da se slika pretvori u sive tonove što može smetati detekciji točnih rubova zbog sličnosti boja. Na slici 4.13. b) se vidi da je smeđa boja slična crvenoj u sivim tonovima, što bi možda uzrokovalo smetnje mreži tijekom detekcije rubova koje bi inače bilo lakše prebroditi da postoje sva tri originalna kanala boje.



**Slika 4.13.** Prikaz koncepta smetnje crvenih objekata nad smeđom pozadinom prilikom pretvaranja slike u sive tonove, a) originalna slika LEGO kocke 3001 na smeđoj pozadini, b) slika pretvorena u sive tonove

Uzevši sve točke dosad navedene u obzir, može se reći da je donekle uspješno proveden prijenos praznina među domenama (engl. *translate the domain gap*) između dobivenih sintetičkih podataka nad kojima se mreža trenirala i realnim uslikanim podacima. Uzevši u obzir da je jedino realno u trening podacima isječene pozadine, a i one su nasumično odabrane i zamućene što može smetati generalizaciji ako par slika zaredom dobije sličnu pozadinu (iako na 20 slika od preko 1000x1000 rezolucije ta vjerojatnost je malena). Dobivena mreža daje zadovoljavajuće rezultate nad stvarnim podacima kad nije nikada bila trenirana nad stvarnim

podacima. Od svih uslikanih slika koje se nalaze u *test\_images* folderu na github repozitoriju, tek 40-50% uspije zadovoljavajuće predvidjeti. Sljedećih 10% slika nezadovoljavajuće predviđa gdje je dobro predvidjelo klasu modela ali sa niskom vjerojatnošću ( $<50\%$ ). Ostatak slika je uporno loše predviđalo kao LEGO 2x4 kocke 3001 ili LEGO kocka kormilo 2340. Moguće je da predviđa kormilo zbog velike površine koju sadrži na strani, pa kad mreža vidi sličnu površinu na danoj slici, aktiviraju se pripadajući neuroni za kormilo.

## 5. ZAKLJUČAK

Korištenje dubokih umjetnih neuronskih mreža je učinkovit način za kreiranje klasifikatora nad slikama, gdje važnost objekta kojeg promatramo leži u njenom oštrom obliku i pravilnim teksturama kao kod LEGO kocaka. Ova spoznaja se može primijeniti i kod drugih sličnih objekata, kao što su kartice pakovanja za lijekove (aluminij na pozadini pakovanja sadrži određenu teksturu i na sebi ima isprintan naziv lijeka i dozu u miligramima, nad time se može duboka mreža učiti).

Za korištenje nad LEGO kockama ispostavlja se da je izvrstan izbor s puno prilagodljivosti s raznim parametrima kojim se može rukovati da se izvede sustav po želji. Primjerice, može se kreirati sustav koji ima više od odabranih 25 klasa ovog rada, kako mogu ostale mreže imati 1000 klasa, može i ova biti trenirana nad više od 25 početno odabranih klasa kao što je Westova mreža.

Za svrhe ovog rada gdje je konačno odabrana duboka umjetna neuronska mreža zasnovana nad prijenosnim učenjem od VGG16 mreže kao trenutno rješenje, mreža je dala relativno brzo treniranje za dostupan hardver i ispunila je očekivane rezultate sa umjereno kvalitetnim prepoznavanjem danih slika LEGO kocaka, nailazi na poteškoće kod teško zamućenih slika.

Postoje moguća poboljšanja sustava:

- kompliciranije mreže se mogu koristiti ako se koristi jači stroj za treniranje, za ovaj rad je korištena grafička kartica AMD RX580 4GB i procesor AMD FX-8350 sa sustavnom memorijom od 8GB DDR3. Optimalno bi bilo s više jačih NVIDIA grafičkih kartica većeg VRAM kapaciteta i procesor istog ranga ili jači s više sustavne memorije, barem 16GB DDR4 ili DDR5
- korištenje mreže koja ima više ulaza, primjerice da sadrži dva ulaza (kao dvije kamere) gdje drugi ulaz prima sliku istog modela ali iz različitog kuta (npr. 90° odmaknuta po nekoj osi da uslika stranu modela) i time uklanja problem mreže ovog rada te postiže mogućnost dobivanja performansi od više nego 95% točnosti (dok s jednim ulazom s pokojom greškom se dobiva maksimalno 85-90% nad sintetičkim podacima)
- alternativa prethodnog poboljšanja je da se isti objekt snima, npr. mobilnom aplikacijom koja detektira model kocke na kameri tako da za svaki model uzme više „uzoraka“ te nad njima čini predviđanja. Na kraju se nad predviđanjima glasa koja je klasa točna.

Doprinos ovog rada bi bilo znanje da postoji isplata za ulaganje vremena u istraživanju podataka (u ovom slučaju istraživanje značajki LEGO kocaka, njihovim mogućim problemima i rješenja za te probleme). Može se prilagoditi duboka umjetna neuronska mreža za korist klasifikacije sličnih objekata, kao npr. druge igračke, elektroničke komponente (prepoznavanje vrijednosti otpornika, modela integriranih krugova, itd.), dijelova za sastavljanje u automatiziranoj strojarnici i slično. Za ovaj rad postoji već fizičko rješenje u obliku Jacquesseovog i Westovog stroja za sortiranje te na Kaggle stranici sa skupom podataka Hazelzeta postoji mnoštvo korisnika koji su isprobali i uspješno kreirali neku vrstu mreže za prepoznavanje LEGO kocaka (nepoznate su performanse nad realnim slikama).

Koristeći jednostavne korake i bolji hardver, svakome je moguće da se uputi u klasifikaciju slika za neki željeni objekt koristeći programska rješenja koje je znanstvena zajednica razvijala zadnjih 20 godina u obliku Pythona, Tensorflowa i Kerasa.

Konačno, može se reći da ovakvo istraživanje nad specifičnim objektima ima smisla raditi jer je moguće dobiti uvid u unutrašnje funkcioniranje dubokih mreža za neke specifične modele promatranja (npr. LEGO kocke), što može pomoći kod općenitog razumijevanja ostalih vrsta dubokih mreža i pri razumijevanju procesa treniranja dubokih mreža.

## PRIZNANJA

Ovaj rad nije sponzoriran sa strane tvrtke LEGO™, niti je primio dopuštenje od LEGO™ za korištenje njihovih modela kocaka.

Po pravilima poštene igre (engl. *fair play*) LEGO™ zahtjeva da se na neautoriziranim radovima ne koristi LEGO™ logo i da se riječ „LEGO“ uvijek koristi kao pridjev (npr. „napravljeno od LEGO kocaka“, ne „napravljeno od LEGO-a“).

## LITERATURA

- [1] L. Vu-Quoc, „File:Neuron3.png“, Wikipedia, 16. rujan 2018., dostupno na: <https://commons.wikimedia.org/wiki/File:Neuron3.png>
- [2] D. Fumo, „A Gentle Introduction To Neural Networks Series — Part 1“, towardsdatascience.com, 4. kolovoza 2017., dostupno na: <https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc>
- [3] K. Hinkelmann, „Neural Networks“, uncam, dostupno na: [http://didattica.cs.unicam.it/lib/exe/fetch.php?media=didattica:magistrale:kebi:ay\\_1718:ke-11\\_neural\\_networks.pdf](http://didattica.cs.unicam.it/lib/exe/fetch.php?media=didattica:magistrale:kebi:ay_1718:ke-11_neural_networks.pdf)
- [4] A. Mikołajczyk, „Data augmentation for improving deep learning in image classification problem“, Research Gate, svibanj 2018., dostupno na: [https://www.researchgate.net/profile/Agnieszka-Mikolajczyk-3/publication/325920702\\_Data\\_augmentation\\_for\\_improving\\_deep\\_learning\\_in\\_image\\_classification\\_problem/links/5d5d5569458515210257607c/Data-augmentation-for-improving-deep-learning-in-image-classification-problem.pdf](https://www.researchgate.net/profile/Agnieszka-Mikolajczyk-3/publication/325920702_Data_augmentation_for_improving_deep_learning_in_image_classification_problem/links/5d5d5569458515210257607c/Data-augmentation-for-improving-deep-learning-in-image-classification-problem.pdf)
- [5] S. Sankaranarayanan, Y. Balaji, A. Jain, S. N. Lim, R. Chellappa, „Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation“, arxiv.org, 1. travanj 2018., dostupno na: <https://arxiv.org/pdf/1711.06969v2.pdf>
- [6] R. Demush, „A Brief History of Computer Vision (and Convolutional Neural Networks)“, hackernoon.com, 26. veljače 2019., dostupno na: <https://hackernoon.com/a-brief-history-of-computer-vision-and-convolutional-neural-networks-8fe8aacc79f3>
- [7] A. Krizhevsky, „ImageNet Classification with Deep Convolutional Neural Networks“, papers.nips.cc, 2012., dostupno na: <https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [8] Y. Huang, S. Shakya, T. Odeleye, „Comparing the Functionality between Virtual Reality and Mixed Reality for Architecture and Construction Uses“, davidpublisher.com, 2019., dostupno na: <https://www.davidpublisher.com/Public/uploads/Contribute/5d5e0535ad919.pdf>
- [9] J. Redmon, A. Farhadi, „YOLO9000: Better, Faster, Stronger“, arxiv.org, 2018., dostupno na: <https://arxiv.org/pdf/1612.08242.pdf>
- [10] „GluonCV: a Deep Learning Toolkit for Computer Vision“, cv.gluon.ai, 2021., dostupno na: <https://cv.gluon.ai/contents.html>
- [11] Microsoft, „COCO – Common Objects in Context“, cocodataset.org, 2021., dostupno na: <https://cocodataset.org/#home>
- [12] J. Mattheij, „Sorting 2 Tons of Lego, Many Questions, Results“, jacquesmattheij.com, 28. lipnja 2017., dostupno na: <https://jacquesmattheij.com/sorting-lego-many-questions-and-this-is-what-the-result-looks-like/#accuracy>
- [13] D. West, „A high-speed computer vision pipeline for the universal LEGO sorting machine“, towardsdatascience.com, 1. kolovoza 2019., dostupno na: <https://towardsdatascience.com/a-high-speed-computer-vision-pipeline-for-the-universal-lego-sorting-machine-253f5a690ef4>



- [14] D. West, „The WORLD'S FIRST Universal LEGO Sorting Machine“, youtube.com, 3. prosinca 2019., dostupno na: <https://www.youtube.com/watch?v=04JkdHEX3Yk>
- [15] C. Shorten, T. M. Khoshgoftaar, „A survey on Image Data Augmentation for Deep Learning“, link.springer.com, 2019., dostupno na: <https://link.springer.com/content/pdf/10.1186/s40537-019-0197-0.pdf>
- [16] Y. LeCun, C. Cortes, C. J.C. Burges, „THE MNIST DATABASE of handwritten digits“, yann.lecun.com, 1998., dostupno na: <http://yann.lecun.com/exdb/mnist/>
- [17] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, „A Comprehensive Survey on Transfer Learning“, arxiv.org, 7. srpnja 2020., dostupno na: <https://arxiv.org/pdf/1911.02685.pdf>
- [18] Z. Wang, Z. Dai, B. Póczos, J. Carbonel, „Characterizing and Avoiding Negative Transfer“, arxiv.org, 5. listopada 2019., dostupno na: <https://arxiv.org/pdf/1811.09751.pdf>
- [19] J. Hazelzet, „Images of LEGO Bricks“: „40,000 images of 50 different LEGO bricks“, kaggle.com, 31. prosinca 2019., dostupno na: <https://www.kaggle.com/joosthazelzet/lego-brick-images>
- [20] Blender zaklada, „Open source 3D creation. Free to use for any purpose, forever.“, blender.org, 2021., dostupno na: <https://www.blender.org/>
- [21] C. Dilmegani, „The Ultimate Guide to Synthetic Data in 2021“, research.aimultiple.com, 19. srpnja 2018., dostupno na: <https://research.aimultiple.com/synthetic-data/>
- [22] stpete\_ishii, „Lego Brick Classify DenseNet201“, kaggle.com, 11. srpanj 2021., dostupno na: <https://www.kaggle.com/stpeteishii/lego-brick-classify-densenet201/notebook>
- [23] Datalira, „Classify Bricks: Compare Transfer Learning Model“, kaggle.com, 3. lipnja 2021., dostupno na: <https://www.kaggle.com/databeru/classify-bricks-compare-transfer-learning-model#4.-Train-the-architecture-with-the-best-result>
- [24] D. West, „How I created over 100,000 labeled LEGO training images“, towardsdatascience.com, 1. ožujka 2019., dostupno na: <https://towardsdatascience.com/how-i-created-over-100-000-labeled-lego-training-images-ec74191bb4ef>
- [25] J. Hazelzet, „One versus two camera setup to recognize LEGO“, kaggle.com, 3. siječnja 2020., dostupno na: <https://www.kaggle.com/joosthazelzet/one-versus-two-camera-setup-to-recognize-lego/log#Conclusion>
- [26] K. Simonyan, A. Zisserman, „VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION“, arxiv.org, 10. travnja 2015., dostupno na: [https://arxiv.org/pdf/1409.1556.pdf\(2014.pdf](https://arxiv.org/pdf/1409.1556.pdf(2014.pdf)
- [27] Y. LeCun, L. Bottou, G. B. Orr, K-R. Müller, „Efficient BackProp“, yann.lecun.com, 1998., dostupno na: <http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf>
- [28] D. P. Kingma, J. L. Ba, „ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION“, arxiv.org, 30. siječnja 2017., dostupno na: <https://arxiv.org/pdf/1412.6980.pdf>

- [29] G. Hinton, N. Srivastava, K. Swersky, „Lecture 6a Overview of mini-batch gradient descent“, cs.toronto.edu, dostupno na:  
<http://www.cs.toronto.edu/~hinton/coursera/lecture6/lec6.pdf>
- [30] J. Duchi, E. Hazan, Y. Singer, „Adaptive Subgradient Methods for Online Learning and Stochastic Optimization“, jmlr.org, 2011., dostupno na:  
<https://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf>
- [31] M. Tripathi, „Underfitting and Overfitting in Machine Learning“, datascience.foundation, 13. srpnja 2020., dostupno na:  
<https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning>
- [32] N. Sristava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, „Dropout: A Simple Way to Prevent Neural Networks from Overfitting“, jmlr.org, lipnja 2014., dostupno na: <https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>
- [33] S. V. Stehman, „Selecting and interpreting measures of thematic classification accuracy“, sciencedirect.com, 28. lipnja 1999., dostupno na:  
[https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
- [34] A. Singh, P. Singh, „Image Classification: A Survey“, a2zjournals.com, 19. prosinca 2020., dostupno na:  
[https://a2zjournals.com/jieee/uploadpdf/Ankita\\_Singh\\_1606489012P49.pdf](https://a2zjournals.com/jieee/uploadpdf/Ankita_Singh_1606489012P49.pdf)
- [35] A. Krizhevsky, V. Nair, G. Hinton, „The CIFAR-10 dataset“, cs.toronto.edu, 2009., dostupno na: <https://www.cs.toronto.edu/~kriz/cifar.html>

## SAŽETAK

Korištenje dubokih umjetnih neuronskih mreža je postao de-facto standard kao prvi korak pri pokušavanju rješavanja nekog problema vezanog za prepoznavanje slika. Ovakvim načinom se dobije znatna količina znanja koju možemo upotrijebiti za analizu korištenih podataka i za daljnje korake potrebne za poboljšavanje korištene mreže (npr. promijene u proširenju skupa podataka ako je preveliko, tipa, smicanje).

U ovom radu su opisane osnove o strojnom učenju i istraženo je kreiranje raznih mreža nad skupom podataka generiranih u Blender-u. Na kraju je odabrana duboka umjetna neuronska mreža temeljena na prijenosnom učenju od VGG16 mreže. Ta mreža je postigla najbolje rezultate (sitno više nego VGG19 temeljena mreža) i po njoj su istraženi ostali parametri kao što su matrica zabune i njene slike značajke, po prvoj metodi je određeno na kojim modelima mreža ima problema dok po drugoj metodi se otkriva kako mreža vidi dane slike u konvolucijskih blokovima.

Ključne riječi: umjetne neuronske mreže, strojno učenje, LEGO™, proširenje skupa podataka, prijenosno učenje

## ABSTRACT

### **LEGO brick classification using neural networks**

The usage of artificial neural networks has become the de-facto standard as the first step in trying to solve a problem related to image recognition. This way one can obtain knowledge about the problem and use it for further analysis concerning the solution (i.e., introducing changes in the data augmentation steps to see how the deep network reacts).

This work explains the basics of machine learning, and it explores the creation of various artificial neural networks which were then trained over an image data set generated in Blender. In the end, the deep artificial neural network based on transfer learning from VGG16 was chosen for its superior characteristics (although the VGG19 based one was close). The network was further explored for its properties, such as its confusion matrix and feature maps, the first research method reveals problematic classes which the network might fail to handle, and the second research method reveals how the network sees the given pictures as the pictures travel down the block of convolution layers.

Key words: artificial neural networks, machine learning, LEGO™, data augmentation, transfer learning

## ŽIVOTOPIS







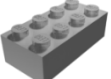
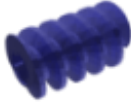
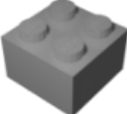
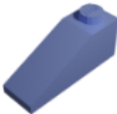
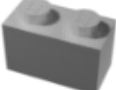





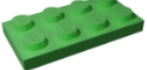



Anto Tufeković je rođen u Sisku 26.01.1998. Početkom je živio u Hrastovcu koji se nalazi u Bjelovarsko-bilogorskoj županiji, ali kasnije se preselio sa obitelji u Tenji u Osječko-baranjskoj županiji. Pohađao je osnovnu školu „Ginko“ u Tenji pa srednju školu „Elektrotehnička i prometna škola Osijek“ i završava sa vrlo dobrim uspjehom.

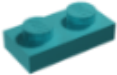




2016. godine upisuje se u „Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek“ (FERIT). 2019. godine završava smjer preddiplomski računarstvo sa završnim radom „Automat za crtanje po školskoj ploči“ kojeg je predstavio na LABUS sajmu iste godine i upisuje na istom fakultetu diplomski smjer „Informacijske i podatkovne znanosti“.

---

Anto Tufeković

## PRILOG A: Popis korištenih modela i njihovi prikazi

|                            |   |   |   |
|----------------------------|---|---|---|
| 2340 Rudder 1x4x3          |    | 3040 Roof tile 1x2<br>45°               |    |
| 2357 Brick corner<br>1x2x2 |    | 3298 Roof tile 2x3<br>33°               |    |
| 2420 Plate corner<br>2x2   |    | 4083 Hanger 1x4x2                       |    |
| 3001 Brick 2x4             |    | 4276 Worm                               |    |
| 3003 Brick 2x2             |   | 4286 roof tile 1x3<br>33°               |   |
| 3004 Brick 1x2             |  | 4864 Wall element<br>1x2x2              |  |
| 3005 Brick 1x1             |  | 6143 Brick D16<br>with cross            |  |
| 3010 Brick 1x4             |  | 6632 Lever 3M                           |  |
| 3020 Plate 2x4             |  | 18575 Double<br>conical wheel Z20<br>1M |  |
| 3022 Plate 2x2             |  | 32140 Technic<br>angled beam 4x2<br>90° |  |

|                    |   |                                   |   |
|--------------------|---|-----------------------------------|---|
| 3023 Plate 1x2     |  | 41678 Cross block<br>fork 2x2     |  |
| 3024 Plate 1x1     |  | 99301 Roof tile<br>inside 3x3 33° |  |
| 3037 Roof tile 2x4 |  |                                   |   |