**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKUFAKULTET ELEKTROTEHNIKE, RAČUNARSTVA** IINFORMACIJSKIH TEHNOLOGIJA **OSIJEK**

**Sveučilišni studij**

Primjena neuronskih mreža u klasificiranju LEGO kocaka

**Diplomski rad**

**Anto Tufeković**

**Osijek, 2021.**

SADRŽAJ

[1. UVOD 1](#_Toc80806764)

[1.1. Ukratko o umjetnim neuronskim mrežama, klasifikaciji i problemu prikupljanja podataka 2](#_Toc80806765)

[1.2. Odabran skup podataka 4](#_Toc80806766)

[1.3. Pregled rada 5](#_Toc80806767)

[1.4. Zadatak završnog rada [[potrebno?]] 5](#_Toc80806768)

[2. Pregled područja rada 6](#_Toc80806769)

[1. ZAKLJUČAK 9](#_Toc80806770)

[PRIZNANJA 11](#_Toc80806771)

[LITERATURA 12](#_Toc80806772)

[SAŽETAK 13](#_Toc80806773)

[ABSTRACT 14](#_Toc80806774)

[PRILOG A: Popis korištenih modela i njihovi prikazi 15](#_Toc80806775)

# UVOD

Često se moraju odrediti točni detalji o nekom objektu ili stvorenju, primjerice psa ili mačke gdje vlasnik se ne zna dovoljno o pasminama da točno utvrdi 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 (naprimjer 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 (naprimjer 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 osobe, osoba sa poteškoćama vida i osjetila dodira (svi formalni lijekovi sadrže ime u Brailleovom pismu) može iskoristiti ovakve tehnologije za npr. prepoznavanje kutije lijekova po njenim karakteristikama (veličina kutije, boja kutije, šare po kutiji, prepoznavanje teksta na kutiji) da odabere 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 sa više klasa, engl. *multiclass classification*). Potrebno je napraviti pregled problema na koje se naiđe tijekom treniranja takve umjetne neuronske mreže, te postaviti ili izvesti rješenja problema.

## Ukratko o umjetnim neuronskim mrežama, klasifikaciji i problemu prikupljanja podataka

U ovom radu će se koristiti **duboke umjetne neuronske mreže** (engl. *deep artificial neural networks*) za potrebe korištenja u algoritmu klasifikatora. Umjetne neuronske mreže su vrsta algoritma koji su bazirani nad neuronskim mrežama životinja, te njihova osnovna građevinska jedinica (neuron) ispada vrlo sličan biološkom neuronu kao na slici 1.1.



**Slika 1.1.** Uzor neurona je prava biološka živčana stanica [[izvor](https://commons.wikimedia.org/wiki/File:Neuron3.png) slike]

Na slici 1.2. se vidi da svaki neuron po uzoru akson – sinapsa – dendrit (izlaz – prijenos – ulaz) ima svoje ulaze () na dendritima, neka funkcija stanice u tijelu, te onda preko aksona ima izlaz() kojim se dalje propagira do kraja sustava. Prethodno spomenuta funkcija stanice neurona određuje kakav će izlaz biti na svakom izlazu ovisno o ulazu, funkcija sadrži u sebi težine koje se pri treningu mijenjaju ovisno o stimulusu te onda odlazi u aktivacijsku funkciju. Najjednostavnije objašnjenje je da svaki izlaz neurona je realna funkcija, te težina pojačava ili oslabljuje „signal“ na danom izlazu.



**Slika 1.2.** Matematički model neurona[[izvor slike](https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc)]

Osnovni matematički model neurona je da svaki ulaz ima svoju težinu [[izvor za ovu sekciju](http://didattica.cs.unicam.it/lib/exe/fetch.php?media=didattica:magistrale:kebi:ay_1718:ke-11_neural_networks.pdf)], težine se postavljaju tijekom treninga ovisno o važnosti ulaza. Na kraju neuron izvršava neku funkciju (često zvana aktivacijska funkcija) nad sumom svih ulaza množeni svojim težinama te to predstavlja kao izlaz svim sljedećim spojenim neuronima.



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

Kao što je prethodno spomenuto, neuroni su često poslagani u slojevima zbog jednostavnosti kreiranja modela, 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 veći uspjeh te glavni cilj tih slojeva je izbjegavanje mrtvih neurona.

**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 bazirane na brojevima (npr. slike imaju piksele koje sadrže tri kanala za boje, najčešće sa vrijednostima od 0 do 255) ili jednostavno uspoređivati dolazeće podatke sa prijašnje treniranim podatcima tako da uspoređuje udaljenost ili sličnost novih podataka od treniranih.

Algoritam koji obavlja klasifikaciju se naziva klasifikator, ako postoje samo dvije klase za klasificirati onda se metoda zove „binarna klasifikacija“, ako imamo više klasa metoda se zove „klasifikacija s više klasa“ a ako imamo metodu gdje neki izlaz može imati više klasa pridruženo onda se radi o „klasifikaciji s više oznaka“.

Za potrebe treniranja neuronskih mreža može se dogoditi situacija da **ne postoji dovoljno dostupnih podataka nad kojima se može trenirati**, razni su razlozi kao npr. poštivanje privatnosti u medicini (naprimjer rendgenske, mamografske, ultrazvučne slike i slično) zbog čega gubimo pristup vrijednim podatcima za treniranje, a potrebna je velika količina da se može trenirati kvalitetna mreža, pogotovo ako je duboka mreža, što je dublja to je više raznolikosti u trening podatcima potrebna da se kvalitetno trenira bez problema pretreniranja i podtreniranja.

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[[izvor](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) za ovo cijelo podpoglavlje? sam izvor ima svoje izvore, možda izvučem nešto]. Za kreiranje nasumične kopije originala često se koriste jednostavne funkcije kao što su afine transformacije (rotacija, skaliranje, translacija, refleksija, smicanje), transformacije nad histogramom (npr. ujednačenje), te mijenjanje svojstva slike kao što su kontrast, svjetlina ili transformacije kao povećanje oštrine, pomućenje slike i izjednačavanje svjetline.

Razlog za proširenjem skupa podataka je poboljšanje generalizacijskog svojstva mreže. Ako se kompleksna mreža trenira nad malim brojem podataka za normalan broj epoha, dogoditi će se podtreniranje mreže jer mreža nije dovoljno trenirana da uspješno prepoznaje objekt na slici, ali ako se trenira nad malim brojem podataka za više epoha mreža će početi pamtiti ulazne podatke i dogoditi će se pretreniranje što opet smeta generalizaciji, jer pri treniranju će imati visoku točnost, a pri validiranju nisku točnost.

## Odabran skup podataka

Ovaj rad je usredotočen na prepoznavanje slika LEGO kocaka. Svaki model LEGO kocke se postavlja kao klasa (npr. 3001 model i 3003 model su svoje klase) te kad bi se ciljali svi modeli svih setova ikad izdanih, postojalo bi bezbroj klasa. Za potrebe ovog rada te za potrebe iskazivanja mogućih problema perspektive uzima se u obzir 25 odabranih modela kao naše klase. Svi modeli i njihovi brojevi modela i prikazi se nalaze u prilogu A.

Odabrane su LEGO kocke jer one sadržavaju jednostavne značajke nad kojima bi se duboke neuronske mreže trebale moći trenirati. Treniranje se svodi na sintetički generiranim podatcima, te tu se uvodi problem generalizacije i razlika u domenama realnih i sintetičkih podataka.

## Pregled rada

Rad sadrži 3 glavna poglavlja i zaključak. Prvo glavno poglavlje se bavi opisom trenutačnog stanja aktualnih radova (engl. *state of the art*) gdje se opisuju neka postojeća rješenja za LEGO klasifikaciju, te dodatni radovi znanstvene zajednice. Drugo glavno poglavlje se bavi generiranjem podataka za ovaj rad te pojedinostima vezano za to [[usporedbom sa drugim postojećim skupovima podataka LEGO kocaka?]]. Treće glavno poglavlje se bavi sa implementiranjem dubokih umjetnih neuronskih mreža koristeći pripadajuće alate (Blender, Python, itd.). U zaključku se nalaze izlazne riječi autora sa mišljenjima o radu, rezultatima rada, doprinosu te prednostima i manama mreže ovog rada.

## Zadatak završnog rada [[potrebno?]]

Proučiti i opisati problem klasificiranja LEGO kocaka. Predložiti model i dizajnirati sustav za klasificiranje te ga ispitati na odgovarajućem podatkovnom skupu.

# Pregled područja rada

Klasificiranje slika su ljudi izvodili već duže vrijeme, AlexNet[[izvor](https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)] je jedan primjer neuronske mreže koja je 2012. godine u ImageNet natjecanju široko pojasnog prepoznavanja postigla prvo mjesto, pobijedivši drugo mjesto za više od 10% manje grešaka u 5 najvećih kategorija koje predstavljaju greške[[top-5 error prevesti ljepše?]], najviše jer su koristili treniranje preko grafičkih kartica. Ovo je privuklo globalnu pozornost prema koristi područja umjetnog učenja i umjetne inteligencije te ubrzo je krenulo napredovanje tih tehnologija. Dan danas se rezultati tehnologija nalaze u raznim poljima svakodnevnice, kamere koriste tehnologije prepoznavanja lica i aktivnog stabiliziranja videosnimke, doktori sa modelima mogu lakše uočiti bolesti na raznim vrstama snimaka, od EKG do rendgenskih slika te moguće je lakše prevesti pokrete tijela u pokrete objekata virtualne stvarnosti uz pomoć neuronskih mreža (npr. za koristi u arhitekturi[[izvor](https://www.davidpublisher.com/Public/uploads/Contribute/5d5e0535ad919.pdf)]).

Joseph Redmon iz Washington sveučilišta je koautor rada [[izvor](https://arxiv.org/pdf/1612.08242.pdf)] gdje opisuje svoju mrežu YOLO9000, koja u stvarnom vremenu može detektirati i klasificirati objekte sa preko 9000 kategorija klasa. U radu spominje razna poboljšanja za originalni YOLO model za detekciju te time ide dalje sa YOLOv2 kojeg iskušava na skupovima podataka PASCAL VOC [[izvor](https://cv.gluon.ai/build/examples_datasets/pascal_voc.html)] i COCO [[izvor](https://cocodataset.org/#home)]. Na PASCAL VOC model radi na 67 okvira po sekundi i dobija performanse od oko 76.8 mAP, dok na COCO na 40 okvira po sekundi dobija performanse od oko 78.6 mAP.

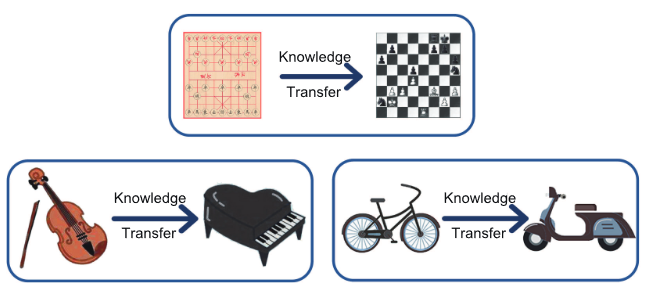
Modificirana VGG16 mreža Jacquesa Mattheija [[izvor](https://jacquesmattheij.com/sorting-lego-many-questions-and-this-is-what-the-result-looks-like/#accuracy)] je mreža koja se koristi da klasificira tip modela LEGO kocke na slici (jeli je dana slika modela kocke, ograde, *Technic* bloka, vegetacije, daje tip, ne direktno model.). Na 60000 primjera podatkovnog skupa za treniranje ima točnost od oko 95%, više za češće modele kao kocke, manje za rjeđe modele 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 sa pokretnom trakom. Stroj polako ubacuje pojedine kocke, koje onda računalo uslika, predvidi mu klasu i koristeći mlaznice potisnutog zraka usmjerava kocke u točnu košaru.

Još jedan primjer stroja je stroj za sortiranje kojeg je izradio Daniel West [[izvor](https://towardsdatascience.com/a-high-speed-computer-vision-pipeline-for-the-universal-lego-sorting-machine-253f5a690ef4)] 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 zajedno svi spreme (slično kao Jacquesov stroj) u zajedničke kante (kocke u kocke, grede sa gredama, komadi figurica sa ostalim komadima, itd.). Uspjeh stroja leži u načinu kako je koncipiran. Originalno je treniran nad sintetičkim podatcima, te kad mu se daju stvarne slike iz pravog svijeta, ima lošu točnost predviđanja. Koristeći nasumično unaprjeđenje skupa podataka, West je prestao pokušavati generirati što sličniju sliku realnome nego je krenuo generirati slike koje se više koncentriraju na specifične detalje modela u pokušaju smanjivanja razlike domena stvarnih i sintetičkih slika. Sa 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, te na tom odabiru mreža se uči. 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, te zbog toga se ne može pravilno uspoređivati sa našom mrežom, iako postoje razni artikli i par videozapisa [[izvor](https://www.youtube.com/watch?v=04JkdHEX3Yk)] koje je West postavio iz čega se vidi efektivnost stroja.

Rad kojeg je napisala Agnieszka Mikołajczyk u Gdansk Sveučilištu tehnologije [[izvor](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)] opisuje razne metode i primjene proširenja skupa podataka za poboljšanje dobivenih 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 sa nasumičnom razinom učinka (npr. translacija se može izvesti samo u rasponu ±20%, smicanje samo ±15°, 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 i pomućenje slike. *Oversampling* se osniva na generiranju sasvim novih podataka učeći uzorke od postojećih ulaznih podataka. Jedna od tih tehnika *oversamplinga* se naziva GAN (engl. *Generative Adversarial Network*). GAN i općenito *oversampling* nisu optimalne za mreže koje trebaju prepoznavati značajke slične LEGO kockama jer one ne mogu pravilno naučiti generirati umjetne podatke modela koji su strogo povezani sa svojim osima (npr. pravokutnike, trokute, kocke) ili strogom pozicijom teksture na modelu (u našem slučaju pozicija dugmadi LEGO kocke na vrhu kocke, ne može 2x2 kocka imati 6 dugmadi). Iz ovog rada se koristi generalno znanje da čak sa malim učinkom proširenja skupa podataka se poboljšavaju rezultati generalizacije dubokih mreža te dodaje motivaciju za istraživanje optimalnog skupa parametara s kojim se mogu postići bolji rezultati.

Dodatan rad koji opisuje proširenje skupa podataka su napisali korisnici *Journal of Big Data* (Dnevnik velikih podataka) Connor Shorten i Taghi M. Khoshgoftaar [[izvor](https://link.springer.com/content/pdf/10.1186/s40537-019-0197-0.pdf)], te slično prethodno navedenom radu opisuju probleme nedostataka ulaznih podataka za treniranje dubokih mreža (npr. u polju medicine) te istražuju moguća rješenja. Isto tako se dijeli na iskrivljavanje podataka i *oversampling*, u kontekstu iskrivljavanja podataka veliku pažnju rad obraća na sigurnost primjene neke transformacije nad ulaznim podatcima (sigurnost u smislu da ulazni podatak ne izgubi svoju konekciju nad oznakom koju nosi, npr. u MNIST[[izvor](http://yann.lecun.com/exdb/mnist/)] skupu podataka za rukom pisane brojeve, važno je ne primjenjivati refleksiju, te držati rotaciju unutar prihvatljivih granica, inače se može dobiti pobuna oko nekih brojeva kao 6 i 9 ako se previše rotira[izvor str. 8]). Još jedan problem koji opisuje je kad su ulazni podatci savršeno centrirani, duboke mreže učene nad takvim podatcima zahtijevaju da su novi podatci za predviđanje isto savršeno centrirani, sličan problem se nalazi u ovom radu jer sve generirane slike su centrirane, pa translacijom se taj problem do neke razine uklanja (npr. imat će i dalje poteškoće prepoznavati modele na samom rubu slike). Još jedna 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, naprimjer mobitelom, neke LEGO kocke. Sama LEGO kocka će ostati oštra ali pozadina neće biti u fokusu te biti će zamućena. U ovom radu se pozadinske slike lagano zamućuju prilikom učenja da bi se simulirao gubitak fokusa pozadine.

Rad Fuzhen Zhuang iz Instituta računalne tehnologije Beijing [[izvor](https://arxiv.org/pdf/1911.02685.pdf)] opisuje prijenosno učenje 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, naprimjer 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 [[izvor](https://arxiv.org/pdf/1911.02685.pdf#figure.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“ [[izvor](https://openaccess.thecvf.com/content_CVPR_2019/papers/Wang_Characterizing_and_Avoiding_Negative_Transfer_CVPR_2019_paper.pdf)], gdje prethodno naučeno znanje mreže smeta učenju nove mreže umjesto da pomogne. 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 baziran nad skupom od korisnika Joost Hazelzet sa stranice Kaggle [[izvor](https://www.kaggle.com/joosthazelzet/lego-brick-images)]) se koristi proširenje skupa podatka koristeći tradicionalne metode (opisane u gore spomenuta dva rada kao afine transformacije), te koristi se prijenosno učenje za svrhe istraživanja tih mogućnosti, što će se kasnije ispostaviti kao najbolje rješenje u ovom radu.

# Implementacija sustava za generiranje podataka

*Blender* [[izvor](https://www.blender.org/)] je program za 3D modeliranje, te 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>

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 sa prefiksom „test“ su direktoriji koji sadrže neuronske mreže nad kojima se treniralo, te 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 rada.

## Generiranje podataka

U Blenderu je potrebno učitati gore spomenutu datoteku scene, unutar nje se nalaze svi potrebni modeli ovog rada, sve što korisnik mora je prenijeti modele iz nevidljive skupine u vidljivu i obratno kako je potrebno. Sintetički podatci za ovaj rad su generirani u Blenderu koristeći skriptu (nalazi se zajedno sa scenom) koja je namještena da uslika zadani model iz raznih kutova. Te slike imaju dva svjetlosna izvora u blizu sredine ali su za jednu prostornu jedinicu odmaknutu 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 te visoko/duboko postavljeno 6 prostornih jedinica).



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

LEGO modeli se skaliraju tako imaju točan odnos veličine međusobno (npr. vizualno pregledati da su na svim modelima dugmadi iste veličine, da su sami modeli iz istog izvora, te time garantirano iste skale, da su tehničke kocke iste dužine (2M, 3M, itd.)) i uslikane su sa jednakom udaljenošću od kamere. LEGO modeli su nasumično obojani od par odabranih boja, kao što su plava, ružičasta, siva, itd., neuronske mreže moraju biti otporne na različite boje u slikama, iako jedan mogući zaobilazak problema je da se slike prije treniranja pretvore u slike sivih tonova.



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

Kamera slika model počevši od skroz gore (slika 3.2.), te u zadanim koracima se kreće sferično oko modela prema dolje. Kako prolazi tako kamera „slika“ model, te generirane slike sprema za daljnje korištenje u treniranju.

Za rad su korišteni modeli koji su ilustrirani u Prilogu A. Cilj je bio koristiti modele koji imaju različite stupnjeve sličnosti, više o tome u podpoglavlju 3.3.

## Usporedba postojećih skupova podataka

Na internetu je uvijek izazov pronaći kvalitetan skup podataka za svrhu nečijeg projekta ili rada. Jedan od izazova spomenutih u podpoglavlju 1.1. u vezi nedostatka podataka za treniranje se veže dalje ovdje. Naime, ako postoji nedostatak potrebno je ili dobiti više podataka iz nekog izvora (npr. više slika iz područja medicine, kao što su slike magnetske rezonancije MRI, podatci dobiveni iz EKG-a pacijenata sa specifičnim bolestima, RTG slike kostiju za svrhe otkrivanja bolesti, itd.) ili je potrebno te podatke „krivotvoriti“ ali u obliku da su što bliže originalnim podatcima u svrsi.

Umjetni podatci imaju široki raspon svrha za koje se mogu koristiti [[izvor](https://research.aimultiple.com/synthetic-data/)] te najčešće pojednostavljuju i pojeftinjuju operacije za koje su vezane, u slučaju ovog rada se ne mora par dana slikati par LEGO kocaka, nego je moguće generirati sve slike u raznim uvjetima, mogu biti sa 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 sa pomućenjem u nekom smjeru, kao da je osoba koja je slikala micala kameru pri slikanju).

Jedan od sintetičkih skupova podataka na internetu je od prethodno spomenutog Joost Hazelzet sa stranice Kaggle[[izvor](https://www.kaggle.com/joosthazelzet/lego-brick-images)\_duplikat]. Njegov skup sadrži 50 klasa što je jednako 50 modela LEGO kocaka, gdje je svaki model uslikan 800 puta iz raznih kutova. Jedan od problema sa Hazelzetovim skupom jest da su sve slike neprozirne. Neprozirnim slikama je teže dodati prozirnost u svrhe generalizacije, te 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 [[izvor](https://www.kaggle.com/stpeteishii/lego-brick-classify-densenet201/notebook)] sa validacijskom točnošću od 66,35% i MobileNet mreža korisnika Datalira na [[izvor](https://www.kaggle.com/databeru/classify-bricks-compare-transfer-learning-model#4.-Train-the-architecture-with-the-best-result)] sa 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 [[izvor](https://towardsdatascience.com/how-i-created-over-100-000-labeled-lego-training-images-ec74191bb4ef)], iako taj skup nije javno objavljen.

## Generirani skup podataka i rezultati

U prethodna dva poglavlja su pojašnjene koristi sintetički generiranih podataka, za slučaj u radu se zna da LEGO kocke imaju oštro definirane oblike (nikada se dvije kocke istog modela ne razlikuju, osim možda po boji), kocke imaju rubove i dugmadi, ploče imaju glatke površine bez dugmadi, štapići za povezivanje su cilindričnog oblika sa specifičnim reljefima na krajevima, kotači su široki cilindri sa reljefima na kontaktnoj površini, itd.

Na gore spomenutoj github stranici rada se nalaze *Jupyter* bilježnice sa modelima, te 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 sintetičkim podatcima jednostavne mreže mogu postići visok rezultat kao što je vidljivo na tablici 3.1. Problem na koji se nalazi 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 trenira mreža sa slikama modela od skroz gore ili dolje, model će naučiti da su to značajke ta dva modela, ali kad dođe u pitanje predviđanje modela takve slike, mreža neće moći sa potpunom sigurnošću reći koji je to model osim ako nije slikano dovoljno u stranu, da se vidi druga strana 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 te tu nije teško da mreža napravi greške, zbog toga nije moguće da mreža postigne neku vrlo visoku točnost. Paralelno tome je odabrano još par modela koji su ili slični ili jedinstveni u izgledu, naprimjer 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 (dali je kocka ili ploča).

a) b)

c) d)

**Slika 3.3.** Prikaz modela 3001 i 3020 za usporedbu poteškoće predviđanja klase modela,  
a) i b) prikazuju pogled ptičjeg oka 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[[izvor](https://www.kaggle.com/joosthazelzet/one-versus-two-camera-setup-to-recognize-lego/log)]. 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

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

# Implementacija ANN za klasifikaciju

U radu se koriste mreže VGG16 i VGG19 [[izvor](https://arxiv.org/pdf/1409.1556.pdf(2014.pdf)\_duplikat] za svrhe prijenosnog učenja. Aktivacijske funkcije umjetnih neurona spomenutih u podpoglavlju 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:

|  |  |
| --- | --- |
|  | (4-1) |

Aktivacijska funkcija mora biti derivabilna zbog svrhe treniranja algoritmom povratnog razmnožavanja (engl. *backpropagation algorithm*) [[izvor](http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf) 9-48]. Na kraju je potrebna optimizacijska funkcija, te odabran je Adam algoritam, koji je oblik algoritma stohastičkog gradijentnog spusta iliti SGD [[izvor](https://arxiv.org/pdf/1412.6980.pdf)]. Adam algoritam je nadogradnja *RMSProp* [[izvor](http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf)] algoritma na takav način da ukomponira svojstva *AdaGrad* [[izvor](https://jmlr.org/papers/volume12/duchi11a/duchi11a.pdf)] algoritma, umjesto da samo bazira stopu učenja za svaki parametar prema srednjoj vrijednosti za svaki parametar, 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 se koristi pri treniranju dubljih neuronskih mreža sa više skrivenih slojeva jer sadrži manju vjerojatnost kreiranja mrtvih neurona zbog prethodno spomenutih učinaka.

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 SGD i povratno razmnožavanje ovaj se postupak iterativno ponavlja dok se ne iskoriste sve mini-šarže u skupu 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, te potrebno je dalje namještati parametre ili ako su zadovoljavajući rezultati spremiti model za daljnje korištenje.

Daljnji problemi pri treniranju se mogu pojaviti u obliku pretreniranja (engl. *overfitting*). Pretreniranje je pojava u kojoj trenirani model postaje previše naviknut na skup podataka na kojem je treniran. To se događa ako je skup podataka za treniranje vrlo malen, ili ako su podatci vrlo slični jedni drugom. Prilikom treniranja mreže potrebno je obratiti pažnju na parametre koje se vraćaju na kraju epohe, 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.



**Slika 4.1.** Grafički prikaz za utjecaj parametara treninga na pretreniranje[[izvor slike](https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning)]

Na slici 4.1. se vide primjeri tri modela sa drugačijim parametrima treniranja. Prvi model je oblika polinomna regresija sa jednim stupnjem, zbog toga se ne može točno podatcima namjestiti (engl. *underfit*). Drugi model ima četiri stupnja, što je dostatno da se model namjesti trening podatcima i da liči originalnoj funkciji. Treći model ima previše stupnjeva te 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 SGD i povratnog razmnožavanja konstantno dolaze slični gradijenti, neuroni će biti naučeni na te specifične podatke, te kad dođu neki drugi dovoljno drugačiji podatci, neuroni neće biti u mogućnosti dati kvalitetan odgovor prema sljedećem sloju, što uzrokuje grešku u npr. predviđanjima i pogoršava izlazne parametre tijekom validacije. Otkrivanje podtreniranja i pretreniranja je jednostavno ako imamo parametre sa treniranja, ako model ima dobre performanse prilikom treniranja i dobre performanse pri validaciji, onda se može reći da je model kvalitetan. Ako model ima loše performanse kod validacije onda je moguće pretreniran, dok loše performanse kod treniranja ukazuju da nije dovoljno treniran, tj. podtreniran je i nedostaje više epoha treniranja.

## Opis slojeva i strukture

Odabrana najbolja mreža po tablici 3.1. je duboka umjetna neuronska mreža bazirana na prijenosnom učenju VGG16 mreže.

# 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, iako 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 sa puno prilagodljivosti sa 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.

Za svrhe ovog rada gdje je konačno odabrana duboka umjetna neuronska mreža bazirana nad prijenosnim učenjem od VGG16 mreže kao trenutno rješenje, mreža je dala relativno brzo treniranje za dostupan hardver, te ispunila je očekivane rezultate za kvalitetnim prepoznavanjem danih slika LEGO kocaka.

Postoje moguća poboljšanja sustava:

* + - * kompliciranije mreže se mogu koristiti ako se koristi jači stroj za treniranje, trenutno je RX580 4GB, AMD FX-8350, 8GB DDR3, optimalno bi bilo sa više jačih NVIDIA grafičkih kartica većeg VRAM kapaciteta, te procesor istog ranga ili jači sa 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 (recimo, 90° odmaknuta po nekoj osi da uslika stranu modela) te time uklanja problem mreže ovog rada i postiže mogućnost dobivanja performansi od više nego 95% točnosti (dok sa jednim ulazom sa pokojom greškom dobiva maksimalno 85-90%)
      * alternativa prethodnog poboljšanja je da se isti objekt snima, naprimjer mobilnom aplikacijom koja detektira model kocke na kameri tako da za svaki model uzme više „uzoraka“ tj. slika te nad njima čini predviđanja. U konačnici se predviđanja mogu pozbrajati te podijeliti sa brojem uzoraka da bi se dobila konačna srednja vjerojatnost, u nadi da usmjerava prema jednom modelu

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, te pri razumijevanju procesa treninga 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, te da se riječ „LEGO“ uvijek koristi kao pridjev (npr. „napravljeno od LEGO kocaka“, ne „napravljeno od LEGO-a“).

# LITERATURA

# 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 te za daljnje korake potrebne za poboljšavanje korištene mreže (naprimjer promijene u proširenju skupa podataka ako je preveliko, tipa, smicanje).

U ovom radu su opisane osnove o strojnom učenju, te 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), te 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, te po drugoj metodi se otkriva kako mreža vidi dane slike u konvolucijskih blokovima.

Ključne riječi: umjetne neuronske mreže, strojno učenje, LEGO™, VGG16, proširenje skupa podataka, prijenosno učenje, matrica zabune, slike značajke

# ABSTRACT

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 (ie. 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™, VGG16, data augmentation, transfer learning, confusion matrix, feature maps

# PRILOG A: Popis korištenih modela i njihovi prikazi

|  |  |  |  |
| --- | --- | --- | --- |
| 2340 Rudder 1x4x3 |  | 3040 Roof tile 1x2 45° |  |
| 2357 Brick corner 1x2x2 |  | 3298 Roof tile 2x3 33° |  |
| 2420 Plate corner 2x2 |  | 4083 Hanger 1x4x2 |  |
| 3001 Brick 2x4 |  | 4276 Worm |  |
| 3003 Brick 2x2 |  | 4286 roof tile 1x3 33° |  |
| 3004 Brick 1x2 |  | 4864 Wall element 1x2x2 |  |
| 3005 Brick 1x1 |  | 6143 Brick D16 with cross |  |
| 3010 Brick 1x4 |  | 6632 Lever 3M |  |
| 3020 Plate 2x4 |  | 18575 Double conical wheel Z20 1M |  |
| 3022 Plate 2x2 |  | 32140 Technic angled beam 4x2 90° |  |
| 3023 Plate 1x2 |  | 41678 Cross block fork 2x2 |  |
| 3024 Plate 1x1 |  | 99301 Roof tile inside 3x3 33° |  |
| 3037 Roof tile 2x4 |  |  |  |